Science

Table of Contents

22 April 2005
Volume 308
Number 5721

NEW THIS WEEK:

RESEARCH

NEWS

News of the Week

News Focus

Products

edited by Stella Hurtley and Phil Szuromi

Funneling Fast Solar Winds

The Sun emits a solar wind that bends and distorts the ionized tails of comets and influences the behavior of Earth's ionosphere. Much of this solar wind consists of an energetic "fast" component whose origin within the solar furnace is not well understood. **Tu *et al.*** (p. 519) used Doppler imagery and magnetic mapping to construct three-dimensional maps of funnels near the surface that are believed to be the source of the fast solar wind. The correlation of ultraviolet emission and magnetic field structure can pinpoint where in the solar atmosphere the fast solar wind is generated.

Microbial Metagenome Analysis

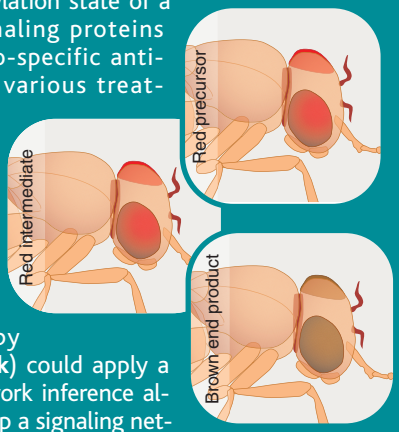
The volume of sequence required to assemble representative whole genomes from complex microbial communities in environmental samples is enormous: up to 100 megabases of sequence is needed to draft a single genome at eightfold coverage, which is feasible for a predominant species, but near impossible for rare species. **Tringe *et al.*** (p. 554) took the alternative strategy of analyzing the gene content of samples from disparate environmental microbial communities. Distinctive metabolic hallmarks indicated selection pressures within the respective habitats. For example, cellobiose phosphorylase was only found in the soil sample but not in the marine samples, and bacteriorhodopsins were found in the surface water samples but none in the deep sea or in soil. The most discriminating operons were for transport of ions and inorganic components. This approach offers a pragmatic and informative route to sifting the enormous volumes of data obtained from metagenome studies.

Emulsions on the Double

Emulsions can be made by mixing one immiscible fluid with another (such as oil and water) to create metastable droplets. Double emulsions, where the core droplet contains smaller droplets of a third fluid, can be more stable but are not easy to prepare in a controlled manner. However, they are of interest to the cosmetic and pharmaceutical industries for delivering a protected liquid product to the user. **Utada *et al.*** (p. 537) have controllably and predictably fabricated double emulsions in a single-step process using a microfluidic device. By injecting fluids in a coaxial geometry, they can keep the fluid reservoirs separate. Droplet sizes are tuned by altering the flow rates.

Modeling Signaling Networks

The prediction of causal influences between components of a signaling network requires detailed modeling from large data sets. Single-cell measurements of the phosphorylation state of a panel of signaling proteins with phospho-specific antibodies after various treatments that influenced cellular signaling provided sufficient data so that **Sachs *et al.*** (p. 523; see the Perspective by **Brent and Lok**) could apply a Bayesian network inference algorithm to map a signaling network and infer causal influences between the components of the network. Known connections were reproduced, and a newly discovered connection was experimentally tested and found indeed to be of biological relevance.



Sharpening Up One's Image

The smallest details that can be imaged are usually limited by diffraction effects on the order of the wavelength of light used for illumination. Recent theoretical work has predicted that it may be possible to overcome the diffraction limit if the properties of the imaging material can be judiciously chosen. In particular, if the electric and magnetic response of the lens material can both be negative, then a flawless image of an object should result. **Fang *et al.*** (p. 534; see the Perspective by **Smith**) used a thin sheet of silver as their superlens and imaged structures with resolution around 1/6 of the wavelength of the illuminating light.

Winds of Wide-Scale Change

Climate warming is affecting atmospheric circulation, ocean circulation, and the marine biological cycle, with implications for weather as well as the global carbon cycle. **Goes *et al.*** (p. 545) provide a striking illustration of how large-scale physical changes can influence biological processes across large areas, even when they are separated by large distances. The decline of winter and spring snow cover in Eurasia that has accompanied mid-latitude warming since 1997 has caused greater continental warming there in the summer. This decline intensified sea-surface winds in the distant western Arabian Sea by creating a steeper thermal gradient. These stronger winds in turn caused intensified upwelling of nutrient-rich water along the Northeast coast of Africa, which increased biological productivity and phytoplankton biomass in the western Arabian Sea.

"Promotin" Signaling by Arrestins

The arrestin proteins got their name because they inhibit signaling from G protein (heterotrimeric guanine nucleotide–binding protein)–coupled receptors like the β2 adrenergic receptor that mediates effects of catecholamines on the heart. However, the proteins β-arrestin 1 and β-arestin 2 have much more versatile roles in signaling. **Lefkowitz and Shenoy** (p. 512) review recent studies showing that β-arrestins also serve as scaffolding proteins that actually enhance signaling by providing binding

CREDITS (TOP TO BOTTOM): BRENT AND LOK; UTADA *ET AL.*

Published by AAAS

sites for other signaling proteins that help produce biological effects of receptor activation. The arrestins even appear to contribute to signaling by structurally distinct receptors, not just G protein–coupled receptors. In their scaffolding role, β-arrestins may transmit activating conformational changes from the receptor to downstream target molecules.

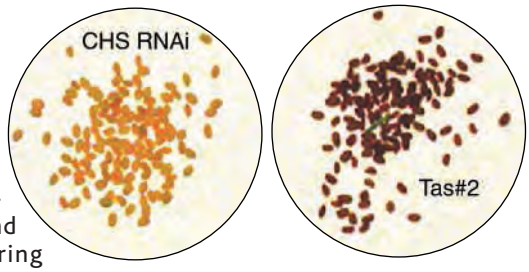
Search and You Will Find

Some of us work in parallel, tackling several tasks at once, and others prefer the serial approach, finishing one task before starting the next. Children searching for a red truck among many toys either examine each object individually (serial) or look first for red objects and for trucks (parallel). **Bichot *et al.*** (p. 529; see the cover and the Perspective by **Wolfe**) now provide evidence that helps to resolve the debate over which approach better describes how visual search operates, in neural terms. A feature-based mechanism (red, truck) operates in a top-down fashion so as to enhance the responsiveness and the synchrony of visual neurons that select for these features. Thus, red toys and trucks evoke more neural activity when a child is searching for a red truck than a brown dog. In addition, a spatial mechanism enhances the responsiveness of visual neurons that are selective for the particular place where the child looks, so that elements of both types of searching contribute.

Elucidating a Plant Defense Mechanism

Arabidopsis strains carrying the gene encoding RPS2 are resistant to infection by the bacterium *Pseudomonas syringae*, which introduces the protease effector AvrRpt2 into plant cells during pathogenesis. **Coaker *et al.*** (p. 548, published online 24 February 2005; see the Perspective by **Schulze-Lefert and Bieri**) now show that the plant's own cyclophilin activates the proteolytic activity of the bacterial effector, AvrRpt2. AvrRpt2 then destroys the intermediate target protein (RIN4) in the plant activating the plant's defensive response. It is possible that such folding of bacterial effector proteases by eukaryotic protein factors may be a common mechanism during pathogenesis.

Molecular Arms Race



Many invading viruses and transposons replicate and transpose through RNA intermediates. These intermediates can be detected by the host cell RNA interference machinery in plants and insects and used to generate small interfering RNAs (siRNAs), critical intermediates in silencing, which can then neutralize the invader. **Lecellier *et al.*** (p. 557; see the news story by **Couzin**) now show that mammalian cells can also use the RNA silencing machinery to help neutralize an invading mammalian virus. Curiously, rather than siRNAs derived from the viral genome being the effector molecules that target the invader for silencing, a host microRNA tags the virus. The importance of the pathway in host defense is supported by the presence of a viral protein that can suppress the silencing effect.

You Scratch My Back...

The interaction between "ant-plants" in the genus *Acacia* and ants in the genus *Pseudomyrmex* is a classic example of a specific, coevolved mutualism; the ants feed on extrafloral nectar produced by the plant, and defend the plant against herbivore attack. The chemical mechanisms underlying this relationship remain mysterious. **Heil *et al.*** (p. 560; see the news story by **Pennisi**) now show that the extrafloral nectar produced by Central American ant-acacias to nourish their resident ants is unattractive to generalist ants because it lacks sucrose. The specialized ants, however, feed on sucrose-free nectar, and they exhibit only very low activity of the sucrose-cleaving enzyme, invertase. The lack of sucrose in the nectar results from invertase activity in the secreted nectar itself.

CREDIT: LECELLIER *ET AL.*

NASA Redux

Six weeks ago, I commandeered this space to report confusion in the ranks at the National Aeronautics and Space Administration (NASA). The former administrator, Sean O'Keefe, was on his way to Louisiana State University; no one knew what would happen to the Hubble telescope; and a host of robotic missions were being put on hold because of rising shuttle costs, congressional pork, and the president's new program: Air Mars, with an intermediate stop at the Moon. Near the end of that piece, I urged that the president appoint a new administrator. To my utter amazement, he did so 24 hours later.

The appointee, Michael Griffin, a respected scientist/engineer from Johns Hopkins, gave the scientific community some encouragement in his confirmation hearing on 12 April. He indicated that once the shuttles start flying again, he would consider sending astronauts to service Hubble. That decision may be controversial, inasmuch as it represents a reversal of O'Keefe's announced intention, but Griffin has some political cover in the form of a National Academies recommendation.

The other parts of Griffin's challenge look much more difficult and could test the comfort of his scientific colleagues in the agency. In this week's *Science* (p. 484), Andrew Lawler sets out a thorough account of those problems. Griffin is a strong proponent of robotic missions, and in 2003 he told the House Science Committee about his commitment to scientific research to understand Earth's environment, the solar system, and the cosmos. Yet Lawler's analysis of NASA's budget suggests that Griffin may be forced to make deep cuts in robotic science in order to keep both old and brand-new commitments to major missions involving human flight.

Indeed, cutoff plans for several science probes were already being developed at NASA as Griffin's appointment was announced. Continuation of the Voyager missions was under threat, although no final decision had been made; and the 2006 budget request from the administration included no funds for an additional group of space science projects totaling $21 million. It has become apparent that NASA simply can't or won't cut out the big human missions, and in order to "keep 'em flying," other, mostly robotic, projects are being scuttled.

Especially distressing to many scientists is the loss of support for Earth observing programs, which lack the political clout of media stars like the Mars rovers or Hubble. The National Academies will soon issue a draft decadal plan for Earth sciences—a sorely needed document like those that have helped astronomers and planetary scientists make their wishes known. It will chart an ambitious program for improving our understanding of oceans, climate, and terrestrial geology and ecosystems. But that vision is not matched by NASA's recent decision to delay or cancel virtually every Earth science mission planned for the coming decade and to terminate several orbiting spacecraft next year.

There is also reason for concern about the future of the scientists who do NASA-supported basic research at other institutions. Deep cuts are now in prospect for these extramural grant programs. That amounts to a transfer of funding from academic institutions to the big industrial contractors who build the vehicles: Think of it like Cal Tech and Stanford paying Lockheed Martin. Nor are changes disadvantaging basic research limited to NASA. A similar transition is under way at the Defense Advanced Research Projects Agency (DARPA), the unit in the Department of Defense that formerly supported some of the most imaginative research programs sponsored by any government agency: the Arpanet, for example, which led to the Internet. Now the DARPA budget has been realigned, with an enlarged share for technical development and less for basic research. University computer science budgets are already feeling the fallout.

Bashing the president on his new exploration vision is probably a waste of breath. A more effective approach would be to insist that exploration is what NASA's science is all about, whether studying the oceans, extrasolar stars, or a Mars ravine, and whether it's done by humans or robots. Finding more money will be hard in a domestic discretionary budget squeezed by growing entitlements and the effect of the tax cuts. But the White House and the Congress must recognize that NASA's superb and diverse research programs should benefit from the president's vision rather than pay a price for it. Let's hope that Griffin, who once observed that the competition between robotic and human missions should not become a zero-sum game, will summon that same wisdom and diplomacy to keep the best science at NASA intact and thriving.

Donald Kennedy
Editor-in-Chief

10.1126/science.1113665

CREDIT: RENEE BOUCHARD/NASA

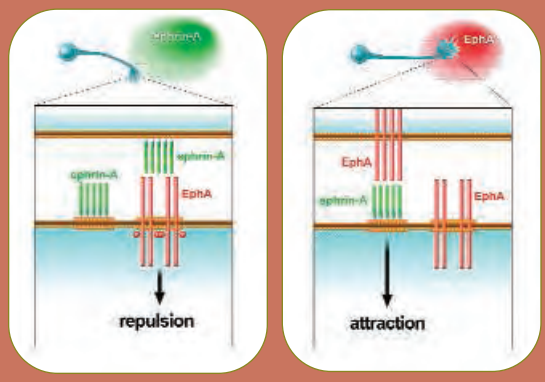
Published by AAAS

edited by Gilbert Chin

NEUROSCIENCE

Stop on Green, Go on Red

Neuronal growth cones flaunt cell surface receptors that sense attractive and repulsive guidance cues as axons make their way to their destinations. Some of these cues are cell-surface proteins, too, and serve as receptor ligands. But what if both receptor and ligand are present in the same growth cone, as is the case with the Eph receptor tyrosine kinases and ephrins, their cognate, membrane-bound ligands?



Repulsion by exogenous ephrin-A (green) and attraction by EphA (red).

Marquardt *et al*. propose that Eph receptors and ephrins segregate into subdomains of the growth cone membrane, allowing them to mediate repulsion and attraction independently. Motor neurons from the chick embryo spinal cord express the receptor EphA and the ligand ephrin-A. When neurons were treated with soluble EphA or ephrin-A, and then with antibodies that promoted clustering, the corresponding cell-surface receptors and ligands were observed to be partitioned into distinct membrane domains on the growth cone. Chimeric EphA and ephrin-A molecules were engineered to force a spatial intermingling of ligand and receptor, and expression of either chimera interfered with the growth cone response to soluble EphA or ephrin-A, indicating that spatial separation of endogenous receptors and ligands facilitates their responses to transcellular cues. The segregation of Eph and ephrin molecules in growth cones may enable axons to see both stop and go signals as they travel to their targets. — LDC

Cell **121**, 127 (2005).

CELL BIOLOGY

Nuclear Waste Disposal

Quality-control systems within the cytosol are important for the overall health of the cell; aberrant proteins (incorrectly assembled or damaged during use) may not function properly, and the cell has mechanisms for disposing of such waste (and recycling the components) if they cannot be repaired. Gardner *et al*. find that a similar quality-control system operates within the cell nucleus of the yeast *Saccharomyces cerevisiae*.

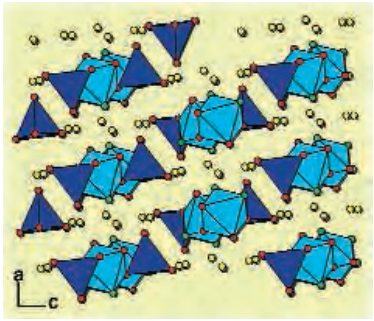
Mutant nuclear proteins are targeted for destruction by the proteasome: a proteolytic machine of the cytosol. A set of nuclear quality-control proteins—San1, which is a ubiquitin-protein ligase, and two ubiquitin-conjugating enzymes, Cdc34 and Ubc1—act together to tag mutant proteins with ubiquitin, the molecular label for proteasomal degradation. San1 possesses a nuclear localization signal that is required for its function, and cells lacking San1 suffer from chronic stress, presumably due to the accumulation of aberrant proteins within the nucleus. Thus, the eukaryotic cell has surveillance and quality-control strategies to protect each of its compartments from the harmful consequences of dysfunctional proteins. — SMH

Cell **120**, 803 (2005).

CHEMISTRY

Cathode Fluoridation

One highly demanding application of lithium batteries is that of charging the capacitors of implantable cardioverter defibrillators. Recharging of the capacitors that deliver the shock to the heart requires batteries that can deliver high voltage and power quickly. For this application, silver vanadium oxide (SVO, $Ag_2V_4O_{11}$) has been a cathode of choice, but there is interest in increasing the capacity that can be delivered above 3 V, which is that portion associated with silver reduction (the rest is associated with V^{5+}/V^{4+} and V^{4+}/V^{3+} couples). Sorensen *et al*. incorporated fluoride into these materials through a low-temperature (150°C) hydrothermal synthesis that yielded $Ag_4V_2O_6F_2$, which increased the silver content of the material without sacrificing the framework structure that allows facile lithium and silver diffusion. This material has about 50% higher capacity above 3 V than does SVO, and because of the fluoride incorporation, delivers it at a potential that is 0.3 V higher. — PDS



Packing in $Ag_4V_2O_6F_2$ (Ag, yellow; O, red; F, green).

J. Am. Chem. Soc. 10.1021/ja050150f (2005).

PSYCHOLOGY

Deciding to Opt In

Humans are social animals, and, as such, it is to be expected that acceptance into a group would confer benefits on oneself, whereas rejection would affect one's behavior adversely. Baumeister *et al*. have performed a set of six experiments to identify the underlying cause of impaired behavior. In this and earlier work, the primary hypothesis has been that social exclusion leads to emotional distress, which in turn has a detrimental impact on task performance. However, in a variety of scenarios, negative mood evoked directly (via bad news) did not affect behavior, and there was no evidence for mood or self-esteem as a mediating factor for the effects of social exclusion on performance. What was observed was a lack of self-regulation, meaning that excluded individuals (in comparison to socially accepted individuals) were less able to drink a healthy but unpleasant-tasting beverage and were more likely to eat unhealthy but tasty snacks. Because the adverse effect of rejection could be ameliorated by introducing a cash incentive for performance, the authors propose that

CREDITS: (TOP) MARQUARDT *ET AL.*, *CELL* **121**, 127 (2005); (BOTTOM) SORENSEN *ET AL.*, *J. AM. CHEM. SOC.* 10.1021/JA050150F (2005)

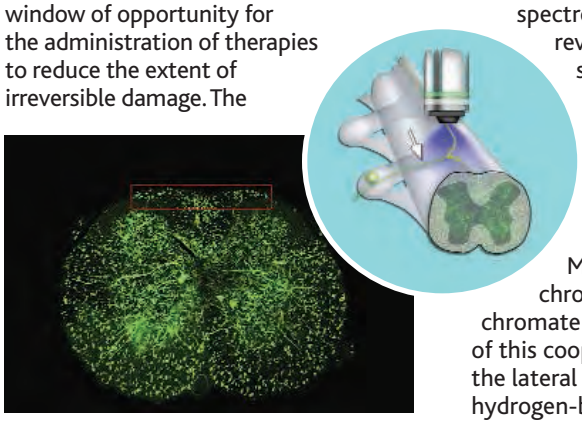
the capacity for self-regulation is intact but that a social rebuff lessens the willingness to make effortful short-term sacrifices in return for longer-term rewards (of good health or a slim physique). Looked at in another way, the consequences of rejection might be reflected at the neural level as a weight that alters the normative balance of decisions when faced with intertemporal choices. — GJC

J. Pers. Soc. Psychol. **88**, 589 (2005).

BIOMEDICINE

Neural Degeneration

When the spinal cord is injured, degeneration of the nerve fibers, or axons, is not instantaneous but rather is believed to occur in several stages over a period of hours. In principle, this delay creates a window of opportunity for the administration of therapies to reduce the extent of irreversible damage. The development of such therapies, however, requires a better understanding of how mammalian axons respond to injury.



GFP-labeled axons (green) in a spinal cord cross section (left) and a schematic of the technique (inset).

Using time-lapse microscopic imaging of living mice expressing green fluorescent protein (GFP) in individual axons, Kerschensteiner *et al.* visualized the axonal response to traumatic injury. Beginning about 20 min after trauma, axons were found to die back at both proximal and distal ends by a rapid and previously uncharacterized fragmentation process termed "acute axonal degeneration." This was followed by slow axonal retraction and ultimately by fragmentation of the axon's distal ends via the well-known Wallerian degeneration. Although many axons mounted a regenerative response within 24 hours of injury, this response was futile because the axons did not grow back to their original targets. This mouse model will likely prove useful for the testing of new therapies for spinal cord injury. — PAK

Nat. Med. 10.1038/nm1229 (2005).

CREDITS: KERSCHENSTEINER *ET AL., NAT. MED.* 10.1038/NM1229 (2005)

ENVIRONMENTAL SCIENCE

Reduced Mobility

Chromate ions (CrO_4^-), such as those in industrial waste streams, are highly toxic, and a better understanding of their transport properties in groundwater would be useful. In the outdoors, the flow of chromate ions can be influenced by myriad chemical and microbial interactions, which researchers usually lump together into measured retardation factors.

Al-Abadleh *et al.* have used a model system to probe the molecular origins and details of retardation in silica-rich soils. They prepared monolayers of carboxylic acid– and ester-terminated alkyl chains, which were attached via siloxanes to a fused quartz substrate, and used second harmonic generation spectroscopy to monitor the reversible binding to these surfaces of aqueous chromate. In comparison to bare silica, the organic acid layers, which are analogous to the humic acids in soil, nearly tripled the retardation factor, whereas esters increased it by 50%. Moreover, the binding energy of chromate to acid increased with chromate concentration, and an analysis of this cooperative behavior quantified the lateral intermolecular forces in a hydrogen-bonded network of acids, perturbed by metal ions. — JSY

J. Phys. Chem. B 10.1021/jp050782o (2005).

ATMOSPHERIC SCIENCE

Fat Coats

It has been suggested that atmospheric aerosols (particles containing a hydrophilic core of sulfate, nitrate, or ammonium salts) may carry organic surfactants on their surface. If so, this would have important effects on the chemical and physical properties of aerosols, as well as consequences for climate and human health. Recent analysis has shown that some marine aerosols do, in fact, sport an outer layer of fatty acids, but whether this is true for other aerosols has been unclear.

Tervahattu *et al.* report that some aerosols of continental origin are coated with fatty acids. They used time-of-flight secondary ion mass spectrometry to detect the presence of these molecules in the outermost 3 nm (of a 0.1- to 1.0-μm particle) in aerosols derived from forest fires and from the burning of coal and straw. — HJS

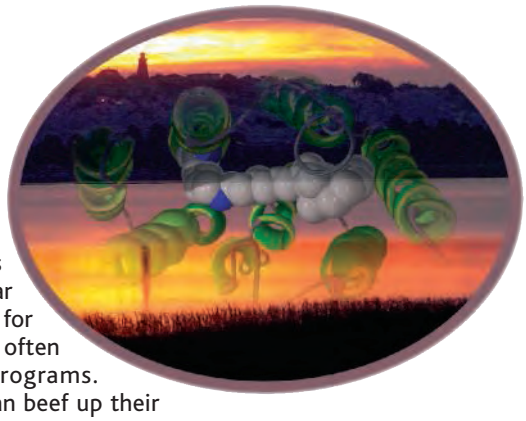
J. Geophys. Res. **110**, 10.1029/2004JD005400 (2005).

NetWatch

edited by Mitch Leslie

EDUCATION

Structural Biology Starter Kit



So-called BLAST searches and fancy 3D molecular graphics may be a snap for veterans, but beginners often need help with the programs. Students and teachers can beef up their structural biology skills at The Molecular Level, a primer from chemist Gale Rhodes of the University of Southern Maine, Portland. Users can bone up on protein structure while learning to use the molecular modeling software DeepView. Another tutorial introduces 10 bioinformatics staples, including the sequence searcher BLAST and the protein analysis tool kit ExPASy. The site offers practice problems, and for the forgetful, there's an organic chemistry refresher.

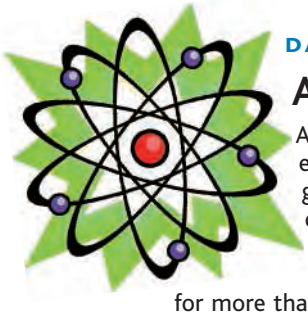
www.usm.maine.edu/~rhodes/index.html

RESOURCES

The Numerical Cell

Looking for a mathematical model of cellular activities, or have you built one you'd like to share? Drop by the new clearinghouse BioModels from a group of organizations including the European Bioinformatics Institute and the SBML Team, an international consortium developing a computer language for describing cell systems. The site stows 20 published models that simulate everything from the conduction of impulses in a neuron to the sugarmaking reactions of photosynthesis. Visitors can download the models in SBML, which is compatible with a host of cell-simulation programs. Annotations spell out the molecules involved, the reactions they participate in, and their cellular locations. Links to databases supply more information about the molecules and reactions.

www.ebi.ac.uk/biomodels



DATABASE

Atomic Almanac

Any periodic table will provide data such as an element's weight and atomic number. But to really get to know, say, molybdenum or strontium, check out the wealth of information at the Elemental Data Index from the National Institute of Standards and Technology (NIST). The site serves as a portal for more than a dozen NIST collections stuffed with atomic measurements, including half-lives and relative abundances for different isotopes and spectroscopy results.

TOOLS

Smarter Searching

Even if you judiciously choose key words and skillfully deploy "ands" and "nots," searching a bibliographic database can return a torrent of hits or skip the paper you want. Researchers looking for an alternative way to bore into the *Caenorhabditis elegans* literature can glide over to Textpresso, a search engine from the operators of the nematode compendium WormBase.

Most bibliographic tools only scan abstracts. But Textpresso digs into the full text of more than 5000 nematode articles, along with some 18,000 abstracts from meetings, the *Worm Breeder's Gazette*, and other sources. And Textpresso lets you narrow your search by categorizing key words and specifying their functions and relationships to other terms. For instance, instead of trawling for all papers on the gene *daf-2*, which governs worm longevity, you can net only publications that record *daf-2* activity in particular types of cells or that identify genes it interacts with. The site includes a similar search engine for the fungus *Neurospora crassa* and prototypes for fruit flies and papers from the *Journal of Neurobiology*. And other teams have launched Textpresso-based libraries for several model organisms, including yeast.

www.textpresso.org

DATABASE

Lives of a Forest

If a tree falls in the moist tropical forest of Panama's Barro Colorado Island, ecologists at the Smithsonian Tropical Research Institute (STRI) might not hear it. But they will find out, thanks to their regular surveys of the locale, which began in 1981. Now anyone can download 20 years' worth of data from this project to monitor tropical trees. About every 5 years, STRI researchers have fanned out through a 50-hectare plot on the island, counting, measuring, and mapping every tree above chest height. The census has tracked more than 350,000 trees from 300 species, including this golden guayacan (*Tabebuia guayacan*; above), and is one of the longest-running ecology studies, says group leader Richard Condit. After filling out a short questionnaire, visitors can download data from the first four surveys and use them to calculate values such as mortality and growth rates for different species.

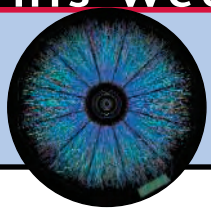
ctfs.si.edu/datasets/bci

Send site suggestions to netwatch@aaas.org. Archive: www.sciencemag.org/netwatch

CREDITS (TOP TO BOTTOM): GALE RHODES; RICHARD CONDIT; CLIPART.COM

Published by AAAS

This Week



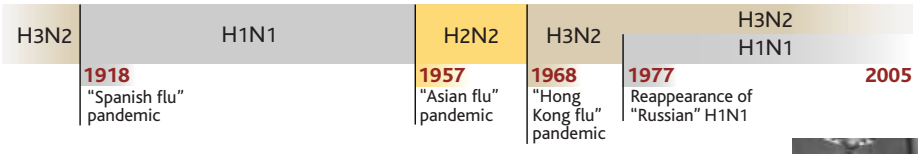
PAGE 479
A quark-gluon plasma in all but name

481
Sugar binds ants to acacia trees

INFLUENZA

Test Kit Error Is Wake-Up Call For 50-Year-Old Foe

Once again, the threat of a flu pandemic made headlines around the world last week, after an influenza A strain called H2N2, which caused the "Asian flu" pandemic of 1957, was accidentally sent to thousands of diagnostic labs. Although experts agree that the episode posed a low risk of a public health catastrophe, it did put an underappreciated question squarely on the agenda: What should the world do with H2N2, a virus not seen in humans since 1968 that is becoming a slightly bigger threat every year?



Jolted into action, the World Health Organization (WHO) says it will soon issue recommendations to bump up safety procedures in labs working with the virus. It will also ask so-called culture collections to remove H2N2 from their catalogs, at least temporarily while the issue is under debate. (The American Type Culture Collection in Manassas, Virginia, already did so "as a precautionary measure last week, a spokesperson said.")

For the long term, WHO plans to reduce the risk of the virus escaping from the thousands of labs storing samples; it might even consider a massive roundup of remaining stocks, akin to the worldwide destruction campaign undertaken after smallpox was eradicated. "It's a peculiar situation," says WHO's principal flu scientist Klaus Stöhr. "We have to ask ourselves: What are we going to do with H2N2 for the next 100, 200 years?"

The kits containing the H2N2 strain were provided by the College of American Pathologists in late 2004 and early 2005 to 3747 labs enrolled in programs that help demonstrate their ability to correctly identify unknown pathogens. It's still unclear why H2N2, and not a current influenza A strain, ended up in the panels, produced by Meridian Bioscience in Cincinnati, Ohio. On Monday, WHO said that the kits had been destroyed in all 18 countries outside the United States that received them; destruction in U.S. labs, which received the vast majority, was expected to be completed shortly.

Although widely described as a "killer strain" in the press, H2N2 was mild as pandemics go when it swept around the globe 48 years ago. The reason that it killed an estimated 1 million to 2 million people, mostly elderly, was not its inherent virulence but because no one had any immunity to it, says Stöhr. At the time, H2N2 completely replaced H1N1, the influenza strain that had burst onto the scene in a much more lethal pandemic in 1918. H2N2 in turn was replaced by H3N2 during the so-called "Hong Kong flu" pandemic of 1968. In 1977, H1N1 reappeared—the result of an escape from the lab, most flu experts think—and since then H3N2 and H1N1 have occurred side by side (see graphic, above). The annual flu vaccine is designed to protect against the most recent versions of both strains and the less fickle influenza B virus.

Researchers don't know what would happen if H2N2 reappeared in the human population right now. No one born after 1968 would have immunity, but the elderly would still have some protection. "I certainly wouldn't expect a full pandemic," says Alan Hampson of the WHO Collaborating Centre for Reference and Research on Influenza in Melbourne, Australia. But H2N2 could become established along with the other circulating strains, he says, complicating vaccine production even more.

Moreover, as each year passes, the risk of a full-blown pandemic rises a notch, says Sylvie van der Werf, a flu researcher at the Pasteur Institute in Paris. Yet in labs around the world, H2N2 is still actively studied under biosafety level (BSL) 2 conditions, a relatively low degree of protection. "People have worked with this for many, many years, not realizing the situation is getting more and more dangerous," says van der Werf.

WHO has drawn up tougher safety recommendations that would require BSL-3 for certain operations with the virus, Stöhr says; they are currently being circulated among experts for comments. But H2N2 is also stored in samples in hundreds or thousands of labs around the world, a problem that WHO discussed internally last year until more urgent issues put it on the back burner. "Now, we're going to reprioritize this," says Stöhr. Countries will be asked to make sure that remaining samples are either destroyed or stored and handled properly.

That effort might eventually evolve into a much larger, more formal exercise to expunge the virus from freezers where it doesn't belong, perhaps supported by a resolution from the World Health Assembly and a verification procedure. Such a process eventually reduced the number of labs holding the smallpox virus to just two in the 1980s; a comparable but less drastic campaign is beginning for poliovirus, the next candidate to be wiped from the planet. But it's not clear that the risk would warrant such a massive operation for H2N2, says Stöhr.

All gone. The emergence of H2N2, known as the Asian flu, caused empty classrooms in 1957. The virus reigned for 11 years before being replaced by H3N2.

It's also not clear how long H2N2 will remain a prisoner. Lab accidents aside, some believe that given the cycling of strains witnessed in the 20th century, nature itself is bound to relaunch H2N2 into the human population at some point. That would make much of the new debate moot.

—MARTIN ENSERINK

CREDIT (RIGHT): © BETTMANN/CORBIS

Published by AAAS

AVIAN INFLUENZA

Outbreak in Northern Vietnam Baffles Experts

World Health Organization (WHO) officials are warily watching an apparent change in the pattern of human infections with the H5N1 avian influenza virus in northern Vietnam. In contrast to the devastatingly high mortality rate of 70% seen previously, the fatality rate in northern Vietnam has plummeted to about 20% since January, according to WHO's office in Hanoi. The cases are occurring in larger clusters—for instance, among five members of one family. In addition, the disease, which has been concentrated among children and young adults, is now afflicting patients of all ages. Such changes suggest that the virus could be evolving to become "less virulent and more infectious," says Peter Cordingley, a spokesperson for WHO's Western Pacific Office in Manila.

WHO officials say there is still no evidence of human-to-human transmission, which could trigger a deadly pandemic. Even the family clusters seem to have been exposed to a common poultry source. "But the pattern of clusters with people getting mildly sick and relatively low mortality is something we haven't seen before in other countries or even in other parts of Vietnam," says Cordingley. However, his WHO colleague in Hanoi, epidemiologist Peter Horby, warns that the pattern could be the result of better surveillance. Viral samples from recent northern Vietnam patients have been sent to the U.S. Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, for comparison to previously recovered samples. Results may be available during the week of 18 April.

Viruses often adapt to their hosts and become less virulent over time. One evolutionary theory, Cordingley explains, is that for the virus to thrive in humans, it can't kill so many of its victims. Although lower mortality may sound reassuring, "even if there is a huge drop in the fatality rate, [a pandemic] would be devastating," warns Scott Dowell of the International Emerging Infections Program, a collaboration of Thailand's Ministry of Health and the U.S. CDC.

Family affairs. Nguyen Si Tuan, 21 (*left*) and his sister, Nguyen Thi Ngoan, 14, are one of a number of family clusters of H5N1 cases in northern Vietnam that have alarmed officials.

International reinforcements are finally arriving in Vietnam; they could help sort out just what is happening. A Canadian team with portable testing labs will bolster the country's own capabilities. And a trio of infectious-disease specialists from the United States, New Zealand, and Australia arrived the week of 11 April to advise the government on public health strategies.

But offers for help on the animal health side are not as forthcoming, says Anton Rychener, the representative of the U.N. Food and Agriculture Organization in Hanoi. One puzzle is why human cases have increased even as outbreaks among poultry have decreased, he says. Another worry is that common chickens may be acquiring resistance, which could enable them to spread the disease asymptomatically. Yet he is unaware of any offers of help from the international community. Technical and financial support is particularly critical to prepare for large-scale poultry vaccination campaigns that might help minimize the chances of humans being exposed to the virus.

–DENNIS NORMILE

U.K. SCIENCE

Industry-Academic Drug Screening Plan Targets CJD

CAMBRIDGE, U.K.—Through mergers and buyouts, GlaxoSmithKline has amassed a huge collection of potential drug compounds and now seems ready to let outsiders glimpse this precious hoard—if the cause is right. Last week the U.K.–based drug giant made what it is calling an "unprecedented" deal to let an academic lab scan its million-plus compounds in hope of finding a treatment for Creutzfeldt-Jacob disease (CJD), the brain-destroying illness caused by prions. The leader of the drug screening project, to be funded initially for 3 years by the U.K.'s Medical Research Council, is prion expert John Collinge of University College London (UCL).

Project leader. UCL's John Collinge.

The plan has been "in the works for some time," says Frank Cooper of Collinge's lab at UCL. But the details are not yet fully worked out. Glaxo spokesperson Gwenan Evans says the company will share data on its compounds and send robotic technicians scurrying through four major U.S. and European facilities to gather up whatever Collinge's lab requests. Evans predicts that the company will turn over a large number of samples. Its capacity is large: "We did over 100 million screens last year," she notes. Glaxo will retain ownership of the compounds, Evans says, but would likely negotiate a no-profit deal if a therapy proved worthwhile.

Others have already started down this path, notably Byron Caughey, a prion researcher at a U.S. National Institutes of Health laboratory in Hamilton, Montana. Caughey says he's screened "thousands" of compounds already but hasn't yet found one that shows much benefit in animals if given after symptoms appear.

CJD and the related "variant CJD" (vCJD), which has been linked to prion-infected beef, are frightening diseases. Death follows soon after the symptoms; deteriorating muscle control and rapid dementia. In Britain, where more than 179,000 cattle in the food chain were confirmed as carrying prion disease in the 1980s and 1990s, vCJD created shock waves. But its toll has been small compared to other diseases: About 150 people have been affected. And that is why the search for a cure is unlikely to get a push from the profit motive. –ELIOT MARSHALL

CREDITS (TOP TO BOTTOM): STR/AP PHOTO; MEDICAL RESEARCH COUNCIL

Published by AAAS

COSMOLOGY

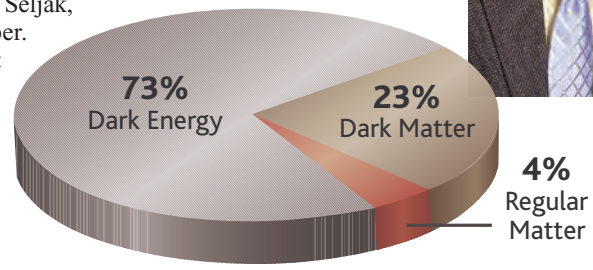
Counterattack Heats Up Dispute Over 'Dark Energy'

Is dark energy an illusion? "I'm not willing to bet my life on it yet," says Edward Kolb, a physicist at the Fermi National Accelerator Laboratory in Batavia, Illinois. "I would bet my collaborators' lives, though." In mid-March, Kolb and three Italian collaborators posted a provocative paper arguing that dark energy—the mysterious antigravity force that makes the universe expand ever faster—is actually a byproduct of enormous ripples in the fabric of spacetime. Kolb's paper created ripples of its own, and now two theorists from Princeton University argue that Kolb's team made an accounting error that invalidates the result.

"They essentially didn't include all the terms of the analysis," says Uroš Seljak, lead author of the antiripple paper. "There's too many papers out there discussing the issue. We thought it was time to make it clear what could and could not be demonstrated."

Kolb's paper, which appeared on the arXiv preprint server (www.arxiv.org), suggested that dark energy—whose effects have been observed by supernova hunters and other astronomers—is not really an energy or a substance. Instead, Kolb says, enormous "perturbations" or ripples in spacetime much larger than the observable universe cause the accelerating expansion of the universe. These ripples, which were caused by the rapid period of inflation just after the big bang, would mimic the fluidlike substance scientists now call dark energy.

Kolb's proposal made headlines and generated a flurry of follow-up papers from physicists around the world. But the attention might be premature, argue Seljak and his colleague, Princeton physicist Chris Hirata. In a paper also posted on the archive, they launched a two-pronged attack on the Kolb hypothesis.



Dark-horse theory. Edward Kolb (*right*) thinks the main ingredient in physicists' recipe for the universe may not exist.

First, using a powerful equation derived from those of general relativity, the two derive a "no-go" theorem that says that huge ripples can't make the universe expand faster and faster. "The equation shows they cannot lead to acceleration," says Seljak. "You cannot have acceleration with only ordinary matter" in the universe; there has to be dark energy.

Next, they attack Kolb's mathematics. Seljak and Hirata argue that in the intricate mathematical calculations leading to the result, Kolb and colleagues inadvertently left out some crucial terms that exactly cancel the effect that they are claiming. "They have been fooled into thinking that there's no cancellation," Seljak says. "These things happen. It's not an easy calculation."

Some physicists, such as Edmund Bertschinger of the Massachusetts Institute of Technology, say Seljak and Hirata have put the matter to rest. "This is definitive in my mind," he says. Kolb, however, holds firm. "I think the no-go theorems eventually will go," he says, adding that he believes Seljak and Hirata have themselves made subtle errors that invalidate their criticisms. "But their work is sharpening our thinking, and we are writing another paper." Whether or not dark energy is making the universe accelerate, the debate over dark energy is itself getting rapidly larger.

—CHARLES SEIFE

HIGH-ENERGY PHYSICS

Latest Data Deal 'Pentaquark' Sightings a Fresh Blow

TAMPA, FLORIDA—The elusive pentaquark may be about to disappear. A new result presented at a meeting* here provides the strongest evidence yet that the much-studied Θ^+ (theta-plus) particle is just a statistical mirage.

"We don't see a structure corresponding to the Θ^+ in this region," says Raffaella De Vita, a physicist at Italy's National Institute for Nuclear Physics in Genova. The new data, from an experiment at the Thomas Jefferson National Laboratory (JLab) in Newport News, Virginia, don't completely rule out the pentaquark, De Vita says. But they do undermine one line of support for the particle's existence and have a much higher statistical significance than the original sightings did. "There's lots of positive results with mediocre statistics, and now one case that can put a nail in it," says Kenneth Hicks, a physicist at Ohio University in Athens who is working on another pentaquark experiment at JLab. "But it's not closing the door just yet."

The pentaquark saga began 2 years ago when a Japanese experiment, SPring-8, seemed to catch a glimpse of a particle, Θ^+, that couldn't be made of two- or three-quark ensembles like all the quarky matter scientists have seen. Within months, other experiments had announced nearly a dozen more sightings of the particle (*Science*, 11 July 2003, p. 153). After data from earlier particle-physics experiments failed to show the Θ^+ or related exotica (*Science*, 19 November 2004, p. 1281), physicists awaited the results from several JLab experiments tailor-made to find the pentaquark.

De Vita revealed the results of the first of those experiments, known as g11. In g11, physicists shined gamma rays at a target full of protons; in theory, a collision between a photon and a proton could create a Θ^+. In 2003, a German collaboration in Bonn using a similar setup claimed to have produced about 60 pentaquarks, a nearly 5-standard-deviation detection. But g11's much more thorough search found nothing. There were huge spikes in the data corresponding to other particles, De Vita says, but none where the Θ^+ should have been.

Another round of JLab results might seal the pentaquark's fate. Hicks says he and colleagues are "very close" to finishing an analysis of data from an experiment that used targets rich in deuterons, atomic nuclei consisting of a proton bound to a neutron. There are theoretical reasons to think deuteron targets might produce Θ^+ particles more readily than proton ones do, says Gerald Miller, a physicist at the University of Washington, Seattle. "If you don't see it in the deuteron, then it's very bad news if you're a pentaquark fan," Miller says.

"I hope the issue will be settled soon," says Curtis Meyer, a physicist at Carnegie Mellon University in Pittsburgh, Pennsylvania. "But I'm not going to buy any pentaquark stock right now."

—CHARLES SEIFE

* American Physical Society April Meeting 2005, 16–19 April.

CREDITS (LEFT TO RIGHT): ADAPTED FROM ROCKY KOLB; FERMILAB

Published by AAAS

HIGH-ENERGY PHYSICS

Unspeakable State of Matter Starts to Reveal Itself—But for How Long?

TAMPA, FLORIDA—Reporters from around the world gathered to hear the announcement, but it didn't come. There was no white smoke, no pronouncement: "Habemus quark-gluon plasma." At a press conference here* on 18 April, scientists working on the Relativistic Heavy Ion Collider (RHIC) celebrated the discovery of a new state of matter not seen since the first moments after the big bang—although they stopped short of claiming to have sighted a long-sought quarry, a quark-gluon plasma. However, the excitement was tempered by a frightening budget profile for nuclear physics and an uncertain future for Brookhaven National Laboratory, the Department of Energy (DOE) facility in Upton, New York, that hosts RHIC.

RHIC, which was conceived in 1983, smashes atoms together at speeds close to the speed of light. Scientists hoped the concentrated energy of the collisions would momentarily melt the protons and neutrons of the nuclei into their constituent quarks and gluons, creating a quark-gluon plasma.

After years of running, RHIC scientists certainly saw something new (*Science*, 20 June 2003, p. 1861; 24 December 2004, p. 2180). Jets of particles flying away from the collision seemed to be moving through a sticky goop rather than through a group of hard protons and neutrons as would ordinarily be the case. Furthermore, after the collision, the system behaved like an expanding puddle of fluid rather than a swarm of particles. "It's an ideal liquid … with essentially no viscosity," says Sam Aronson, Brookhaven's associate laboratory director for high-energy and nuclear physics. "[It is] as perfect a fluid as calculations would allow."

It was also a surprise. "We expected a hot and dense gas," says Aronson. Instead, the unshackled quarks and gluons are interacting with each other much more strongly than anticipated. As a result, the RHIC scientists could not agree to put the label "quark-gluon plasma" on the substance.

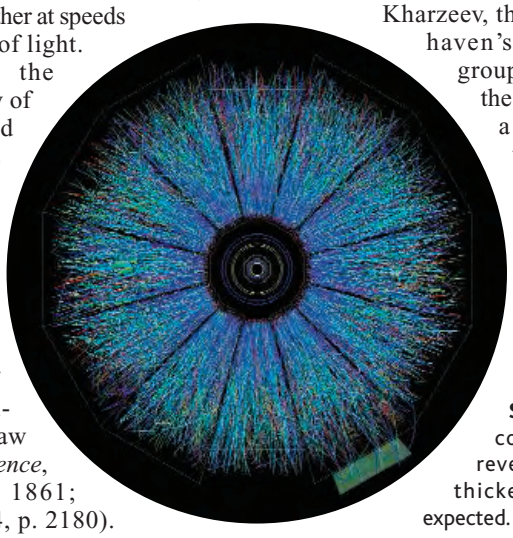
They still can't. "Yes, I think it's a quark-gluon plasma," says Aronson. "The theoretical community—large parts of it—say this is it." Experimental physicists, however, want more direct evidence that the quarks and gluons are roaming completely free. "To some degree, it's a matter of taste," he says. "I think it'll be resolved pretty soon."

Meanwhile, scientists are already figuring out some of the puzzles that the new state of matter has yielded, including a baffling difference in the behavior of two-quark mesons and three-quark baryons created in the collision (*Science*, 25 October 2002, p. 718). "A perfect fluid expands very fast, and the various elements of the fluid have the same velocity," says Dmitri Kharzeev, the leader of Brookhaven's nuclear theory group. "This means that the heavy particles get a bigger boost than the light ones."

But RHIC physicists may not get a chance to study the new state of matter for much longer. DOE's high-energy physics budget request for 2006 cuts funding by more than 8%, and the run time for the RHIC accelerator will drop from 30 weeks to 12 weeks. Aronson says tight budgets will likely force the lab to lay off about 40 people, or 10% of the staff operating RHIC. And a subcommittee of DOE's Nuclear Science Advisory Committee is currently mulling over the future of nuclear physics: With current budget prospects, the department may have to choose between closing Brookhaven and closing the Thomas Jefferson National Accelerator Facility in Newport News, Virginia.

"It's tragic that one has to cut funding for research when such tremendous progress is being made," laments Wit Busza, a physicist at the Massachusetts Institute of Technology who works on one of the four experiments at RHIC.

–CHARLES SEIFE



Surprise. Debris from colliding gold atoms reveals a nuclear puree thicker than physicists expected.

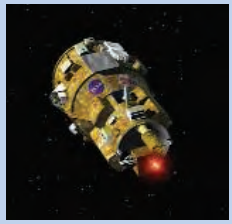
CREDITS (TOP TO BOTTOM): NASA; BROOKHAVEN NATIONAL LABORATORY

* American Physical Society April Meeting 2005, 16–19 April.

ScienceScope

NASA Dart Misses Bull's-Eye

NASA's plan to conduct sophisticated operations in space using robots met with a technical setback last weekend when an agency spacecraft designed to rendezvous automatically with an orbiting satellite shut itself down. The $110 million DART mission—short for Demonstration of Autonomous Rendezvous Technology—was supposed to come within 5 meters of the satellite and execute a series of maneuvers. But as DART came within 100 meters of the satellite, its sensors showed that the NASA probe was using too much fuel and automatically shut off. The probe then put itself into another orbit, where it will degrade and eventually burn up without posing a hazard.



NASA has set up a team to investigate what went wrong with what was intended as a flight demonstrator for future missions. The technology is meant to help NASA deliver cargo to the international space station, service civilian, commercial, and military satellites, and help build larger spacecraft to carry humans to the moon and Mars.

–ANDREW LAWLER

Canadian Climate Plan Silent on Funding

TORONTO—Senior Canadian climate change researchers are fuming at the lack of funding for science in an $8 billion climate change mitigation plan put forward last week by the government. "So far, when it comes to science, Canada's climate change plan is all talk and no action," says Michel Béland of the Meteorological Service of Canada (MSC) here.

Scientists in MSC's atmospheric and climate science unit have been pressing the government for 2 years to renew several soon-to-expire climate change science programs. Last year they recommended increases for earth observation, carbon sinks, ocean sinks, "tipping point" thresholds, and climate change adaptation. But although the new plan acknowledges a need for more science, its focus is on spelling out how the country would meet its commitment under the Kyoto Protocol to reduce greenhouse gas emissions by an estimated 30% by 2012.

–PAUL WEBSTER

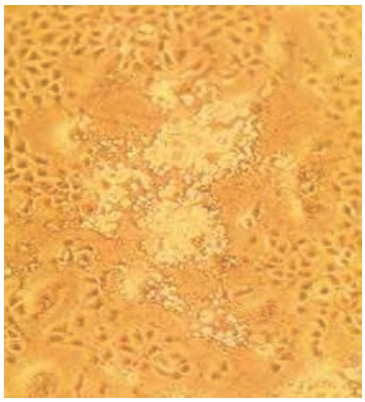
Published by AAAS

MOLECULAR BIOLOGY

Human RNA Slows Down a Primate Retrovirus

A multiyear search has led a 30-something molecular biologist and his colleagues to a new way that human cells fend off viruses. A similar defense system, called RNA silencing because short RNA molecules shut down specific genes, is known to protect plants and insects from viruses, but until now a similar immune mechanism hadn't been detected in mammalian cells.

On page 557, Olivier Voinnet of the CNRS Institute of Plant Molecular Biology in Strasbourg, France, and his team describe a single so-called microRNA, out of the hundreds of different kinds of microRNAs in human cells, that appears to restrict the proliferation of a retrovirus, the viral family to which HIV belongs. Voinnet and others suspect that additional mammalian microRNAs could also have antiviral talents. The work may even explain why it's been tough to identify the roles of so many human microRNAs: They might target the genetic material of foreign viruses, not human genes.



Viral suds. A primate virus creates a foamy residue by killing cells.

"This microRNA may be another arm of immunity," says Shou-Wei Ding of the University of California, Riverside, who reported in 2002 that similar RNAs control viruses infecting fruit flies.

This microRNA is apparently a nonlethal weapon. By shutting down viral genes, it reins in but does not kill a virus, notes David Baulcombe of the Sainsbury Laboratory in Norwich, U.K., who helped identify an RNA-silencing defense system in plants in the late 1990s.

It was graduate work in Baulcombe's lab that drew Voinnet to the challenge of finding an RNA-based antiviral defense in mammals. He selected a primate retrovirus called primate foamy virus type 1 (PFV-1); relatively harmless to humans, it doesn't infect people unless they're bitten by a monkey, Voinnet says.

Working with postdoc Charles Lecellier, Voinnet first used a viral protein called P19 to determine whether human cells even use RNA silencing against viruses. Many plant viruses produce P19 as one way to squelch a plant's RNA silencing system and give themselves the upper hand. Voinnet's team expressed P19 in some human ovarian cancer cells but not in others, then infected all the cells with PFV-1. In the P19-making cells, the virus replicated much more easily, suggesting that the protein had stymied some defense that continued to work in the normal cells.

Here, however, the team's findings diverged from those in plants and insects. In their version of antiviral RNA silencing, plants and insects turn a virus's genetic material back on itself. Many viruses replicate by creating double-stranded RNA copies of their genomes; infected plant or insect cells contain enzymes that chop up this viral RNA into smaller strands that target the genetic material of the virus and destroy it. But to Lecellier and Voinnet's surprise, their team failed to find ▶

HIGHER EDUCATION

Bill Offers Break on Loans to Boost Study of Science

An estimated 100,000 college graduates could save up to $10,000 each under proposed federal legislation to increase the number of U.S. citizens pursuing science and engineering careers.

The bill, introduced last week in both the U.S. House and Senate, would forgive the interest on federal loans for science, technology, engineering, and mathematics (STEM) majors who work in science-related occupations for 5 years after they graduate. If passed, the legislation (H.R. 1547 and S. 765) would create the largest program of its type in the history of U.S. higher education. But some observers say that the amount is too small to steer students into science careers.

Several federal agencies already offer a variety of scholarship and loan forgiveness programs to attract more U.S. citizens into the sciences. Some have paid off, contributing to a 10% increase in the number of domestic students earning STEM degrees from 1991 to 2001, but policymakers say more are needed. Under the Math and Science Incentive Act of 2005, each STEM major would receive a waiver of up to $10,000 in loan interest in exchange for agreeing to work as a science or math teacher or as a STEM professional for five consecutive years. The program would be funded by the Education Department, which already runs a similar program that erases up to $17,500 in student loans in return for 5 years of teaching in an impoverished school district.

"We understand incentives in baseball, basketball, and football," says former House Speaker Newt Gingrich, whose new book *Winning the Future* inspired Representative Frank Wolf (R–VA) to propose the legislation. "There's no reason why incentives won't work in education."

> **"There's no reason why** incentives won't work in education."
>
> —Newt Gingrich

Others are not so sure. Economist Anthony Carnevale, a senior fellow at the nonprofit National Center on Education and the Economy in Washington, D.C., calls the legislation "a symbolic gesture" and doubts it will influence students who aren't already headed in that direction. "Few students make a career decision based on how much interest they might be able to save on a loan 10 years down the road," he says. Naga Kodali, a collections manager for the federal Perkins student loan program at the University of Maryland, College Park, agrees. "If you want to entice students to major in a discipline that requires significant academic effort, you have to offer a comprehensive financial package," he says.

Congress followed a similar line of reasoning in last year's Taxpayer-Teacher Protection Act, which more than tripled the amount forgiven under the Education Department's program to attract qualified math and science teachers to low-income schools. "The feeling was that we needed a better incentive," says Susan Sclafani, assistant secretary of vocational and adult education at the Education Department.

A successful loan-forgiveness program could have the unintended negative effect of crimping the flow of students into graduate school, worries Daryl Chubin, director of the Center for Advancing Science and Engineering Capacity at AAAS, the publisher of *Science*. "Students would presumably enter the workforce immediately after earning the baccalaureate," he says. Forsaking graduate school, he adds, could ultimately put these students at a disadvantage in a competitive job market.

–YUDHIJIT BHATTACHARJEE

CREDIT: C. H. LECELLIER *ET AL.*

Published by AAAS

small RNA molecules derived from PFV-1 in the human cells.

A Spanish friend of Voinnet, Cesar Llave, a plant molecular biologist at the Spanish Research Consul in Madrid, then told him about a recent find: Several plant microRNAs, Llave had just determined, match viral genomes. This left Voinnet wondering whether human cells could directly manufacture small RNAs that thwart PFV-1. Then there would be no need to exploit the virus's own RNA, as plants and insects can. The theory was especially appealing because a single mammalian microRNA might target multiple kinds of viruses.

A comparison of PFV-1's RNA genome and human microRNAs revealed several microRNAs that could potentially silence PFV-1 gene expression. Indeed, when the researchers blocked one of the microRNAs, miRNA-32, the virus nearly doubled its replication rate in cells. Voinnet's group also found that PFV-1 makes a protein called Tas that seems to suppress the microRNA's ability to tackle the virus.

Still, the antiviral potency of miRNA-32 remains unclear. Because PFV-1 infects primates, not humans, researchers need to test whether the virus provokes the same RNA-silencing response in primate cells, says Ben Berkhout, a retrovirologist at the University of Amsterdam in the Netherlands.

–JENNIFER COUZIN

ECOLOGY

Sucrose-Free Sips Suit Acacia Ants

The thorny acacia tree has strong allies: vicious, centimeter-long ants whose nasty bite scares off plant-eating animals and also humans. In return for defending acacias, the ants get free meals and places to live. The key to this sweet deal is the sucrose-free nectar provided by the plant, says Martin Heil, an ecologist at the Max Planck Institute for Chemical Ecology in Jena, Germany. As he and his colleagues report on page 560, a sucrose-degrading enzyme produced by the acacia customizes its nectar to appeal to the right ant partners. The defensive-minded ants that protect the tree prefer their nectar without sucrose, while other ants do not, the researchers found.

Feeding station. In return for fending off the acacia's enemies, ants feast on its nectar-filled globules.

CREDIT: M. HEIL *ET AL.*

Furthermore, the acacia ants have actually decreased their own production of the same sucrose-degrading enzyme, reinforcing this particular pairing of insect with plant. The work "gives one of the first examples of a biochemical basis for behavior difference in plant-insect mutualisms," says Robert Thornburg, a biochemist at Iowa State University in Ames. "It shows that coevolutionary trends can be underlain by biochemistry."

Biologists have documented many cases of coevolution, wherein two species provide for each other and through time develop mutual dependency. Researchers have long known that acacias and *Pseudomyrmex* ants co-mingle: The ants fend off herbivores and in return live inside the safety of *Acacia*'s thorns and eat the plant's nectar. Because each seedling must reestablish this relationship, the plant must have a way to attract the right ants, Heil explains.

Wilhelm Boland, a chemist at the Jena institute and Heil's collaborator, proposed that the key lies in the seedling's extra floral nectar. Working at two sites in Mexico, Heil exposed *Pseudomyrmex* ants and other ants to nectars from four swollen-thorn acacia species and from three other *Acacia* species that don't depend on the specialized ants. He also tested all the ants' preferences for solutions containing varying kinds and amounts of sugar.

All 11 of the ant species that don't live on swollen-thorn acacias bypassed those trees' nectar, whereas the two species of acacia specialists went right to it and, for the most part, rejected the other nectars. The various ants differed in their tastes for the sugar mixtures as well, says Heil. The nonresident ants headed for solutions filled with sucrose, whereas the acacia ants lapped up solutions lacking this particular sugar. When Heil's team added sucrose to the swollen-thorn ▶

ScienceScope

Challenge to Animal Studies

An undercover investigation at Cambridge University by a group that opposes vivisection is prompting a review of how animal experiments are licensed in the United Kingdom. Last week, High Court Judge Stanley Burnton agreed to allow a judicial review of two of the six charges against the U.K. Home Office, the licensing body, raised by the British Union for the Abolition of Vivisection.

U.K. scientists must apply for licenses from the Home Office before conducting research on animals. The union is challenging the Home Office's licensing decisions, prompted by experiments at Cambridge that involved inducing stroke in marmosets to study brain function in stroke and Parkinson's disease sufferers. Among the union's concerns are that the marmosets were deprived of water.

–FIONA PROFFITT

SLAC Plays Catch-Up

Particle physicists at the Stanford Linear Accelerator Center (SLAC) started taking data this week for the first time since an 11 October 2004 electrical accident nearly killed a technician and shut down the particle collider in Menlo Park, California (*Science*, 29 October 2004, p. 788). While SLAC overhauled its safety practices, physicists at SLAC's rival, Japan's KEK laboratory in Tsukuba, kept cranking out data on the differences between matter and antimatter; they now have 40% more data than their SLAC counterparts.

To narrow the gap, SLAC researchers plan to skip this summer's 3-month scheduled downtime. And KEK physicists aren't gloating. "The fact that [SLAC] had this long shutdown is a big setback for the entire field," says Alan Schwartz of the University of Cincinnati, Ohio, who works at the Tsukuba lab.

–ADRIAN CHO

Saying No to Invasives

A bipartisan set of legislators has called for a comprehensive federal effort to protect the country against aquatic invasive species. The bills introduced last week (H.R. 1592, 1593, and S. 507) would authorize a $25 million research program and an interagency council to coordinate federal activities.

The act would also require that ships treat their ballast water to eliminate unwanted organisms, although how they'll do so isn't clear. Treatments using heat or ultraviolet light are being tested, says James Carlton, a marine invasions ecologist at Williams College in Williamstown, Massachusetts.

–AMITABH AVASTHI

Published by AAAS

acacia's nectar, the two groups of ants switched roles.

Heil and his colleagues attribute the sucrose-depleted nectar of the acacia to an enzyme called invertase, which is secreted into the nectar by the plant and breaks down sucrose into glucose and fructose. Invertase activity was 10 times greater in the nectar of the swollen-thorn acacias than in the nectar of plants that don't have ant partners.

"This study reveals that specificity can be achieved relatively simply," says Anurag Agrawal, an ecologist at Cornell University. He predicts that other organisms also home in on the sucrose-poor nectar and coexist with the ant-plant pair. "Though the relationship is specific, it is unlikely to be purely a two-species interaction," says Agrawal.

Diane Davidson, a tropical ecologist at the University of Utah in Salt Lake City, calls the Heil study "rigorous" but wonders if the acacia's ant partners add sucrose-degrading microbes to the nectar. Other strategies could also be used by acacias, she notes. For example, some plants secrete wax that only specialized "wax runner" ants can travel on.

Nonetheless, says Thornburg, Heil and his colleagues "are actually starting to get to the mechanisms" of mutualism. How sweet.

–ELIZABETH PENNISI

CLIMATE CHANGE

Global Warming Skeptic Argues U.S. Position in Suit

The U.S. government has enlisted an outspoken skeptic of global warming in a legal fight with environmental groups over U.S. funding for overseas energy projects. The move has angered several prominent climate researchers, however, who say the government's arguments fly in the face of scientific consensus about both the causes and possible consequences of global warming.

"Impacts ... will include sea level rise, ... disturbances of ecosystems, ... [and] an accelerated reduction of water storage in winter snowpack."

—Michael MacCracken, in brief for plaintiffs

On 29 April, a federal district court in San Francisco will hear a case (*Friends of the Earth* v. *Peter Watson*) about whether the National Environmental Policy Act (NEPA) should apply to projects supported by the Export-Import Bank and the Overseas Private Investment Corporation. The act requires the government to assess actions that could alter the environment. The plaintiffs in the case, which include several environmental groups and four western U.S. municipalities, argue that the federally supported projects—including oil drilling, pipelines, and commercial power plants—contribute to global warming, which in turn affects U.S. economic interests and its citizens. That connection is essential to establish their legal right, or standing, to bring suit.

To counter that claim, the Justice Department argues that "[t]he basic connection between human induced greenhouse gas emissions and observed climate itself has not been established." It buttresses its case with a 41-page statement from David Legates, head of the Center for Climatic Research at the University of Delaware, Newark.

Legates begins by attacking the evidence for the 0.6°C rise in temperature in the 20th century cited by the Intergovernmental Panel on Climate Change (IPCC) in Geneva, Switzerland, in its 2001 report and by the plaintiffs. The proximity of temperature gauges to cities, he says, has artificially elevated reported temperatures. He also points to natural variability as an important factor, citing a 2004 study that suggested solar variability may have contributed up to 0.25°C of the recent warming. As for future impacts, he says surface temperatures in Greenland are falling, coral bleaching is a beneficial response to stress, and the impact of droughts has been relatively benign in the 20th century. Legates says a Canadian climate model that plaintiffs cite to show potential changes in surface temperatures and moisture across North America is "extreme" and "overstated."

The plaintiffs counter with a 45-page brief from climate researcher Michael MacCracken, former head of the Office of the U.S. Global Change Research Program. In an interview, MacCracken called the Legates document "an attempt to go back and reargue the IPCC." Core findings of the IPCC, he says, have been repeatedly confirmed, including the 0.6°C increase in the last century. The urban heat effect has been discounted and cannot explain the warming oceans, says Thomas Wigley, a climatologist at the National Center for Atmospheric Research (NCAR) in Boulder, Colorado. Legates's arguments on solar variability are "standard skeptic crap" that has been discredited, Wigley declares.

MacCracken says Legates's assertion that Greenland is cooling is "wishful thinking," pointing to vast melting around the landmass documented in the recent Arctic Climate Impact Assessment. Severe droughts are on the increase, says IPCC lead author Kevin Trenberth of NCAR. As for Legates's criticism of the Canadian model, MacCracken notes that relevant government agencies have approved the 2000 U.S. National Assessment in which the model was put to use. "It's a selective use of studies and half-truths," Trenberth says about Legates's arguments.

In an interview with *Science*, Legates says he's standing his ground. He questions whether the IPCC represents a true consensus, claiming "a lot of dissenting views." He defends the studies he cites and attacks the Arctic assessment, which he says ignores natural Arctic cycles. Connecting emissions overseas to stateside impacts is simply tenuous, he maintains, adding that the plaintiffs are being selective in choosing the most dire projections.

"Significant questions still remain as to [whether] this ... rise in air temperature can be attributed to anthropogenic increases in greenhouse gas concentrations."

—David Legates, in brief for government

Previous legal attempts to force the government to report carbon dioxide emissions under NEPA, by linking those emissions to climate impacts, have failed. But a 2003 ruling in a suit over natural gas turbines found the failure to disclose CO_2 emissions "counter to NEPA." Earlier this month a federal appeals court heard arguments in a suit that would require the Environmental Protection Agency to regulate CO_2 emitted by motor vehicles.

–ELI KINTISCH

CREDITS (TOP TO BOTTOM): MICHAEL MACCRACKEN; PHOTO COURTESY OF KATHY ATKINSON/UNIVERSITY OF DELAWARE

Published by AAAS

GERMAN SCIENCE

Plan to Boost University Research Caught in Political Crossfire

BERLIN—A bitter dispute over who has responsibility for German universities is blocking a federal government plan to spend nearly €2 billion on cutting-edge research. On 14 April, the latest attempt at compromise ended in disappointment for scientists and university administrators who have been anxiously awaiting the start of a so-called Excellence Initiative designed to boost the fortunes of Germany's most competitive universities, which have suffered decades of tight budgets, aging faculty, and expanding student populations.

Still waiting. The University of Heidelberg is a leading candidate for funding under a stalled program that would support Germany's top universities.

The stakes are high. The proposed initiative would make €1.9 billion ($2.5 billion) available through 2011 under three programs: up to €1 million per year for 40 new graduate schools, €6.5 million yearly for each of 30 "excellence clusters" that would increase cooperation between universities and other research centers, and €21 million a year for 10 universities that develop university-wide strategies to boost themselves to world-class status. The federal government would cover 75% of the program, with state governments covering the rest. An accompanying "pact for research and innovation" would guarantee 3% increases for Germany's nonuniversity research institutes, including the Max Planck Society, through 2010.

The targeted university funding is a dramatic change in Germany, where decades of egalitarian policies have sought to ensure equal access to universities nationwide and "elitism" has been taboo. In January 2004, however, Research and Education Minister Edelgard Bulmahn, a member of the governing Social Democrats, announced that she wanted to fund a program to create a handful of world-class universities that would attract students and researchers from around the globe (*Science*, 11 June 2004, p. 1579).

The German constitution assigns responsibility for universities to the 16 German *Länder*, or states, and several state leaders—chiefly from the opposition Christian Democratic party—protested, saying the plan overstepped the federal government's powers. Months of negotiations produced the three-pronged funding plan, and state and federal leaders have been near agreement at least twice. Most recently, on 6 April, the science ministers from all 16 states agreed to a final proposal, and it looked as though the plan would go forward. A week later, however, on 14 April, Christian Democrat leaders balked and refused to sign off. In particular, the leader of Hessen, Roland Koch, has said the plan would create an unacceptable "two-tier system" among Germany's 99 universities.

The continuing blockade is "completely incomprehensible," says Ernst-Ludwig Winnacker, president of the DFG research funding organization. "A few politicians are … tarnishing the international reputation of German research." Bulmahn said in a press conference a day after the latest breakdown that she is ready for further negotiations and "will continue the fight." However, most observers predict that the stalemate will continue at least until after state elections in North Rhine-Westphalia on 22 May—where the Christian Democrats are hoping for a big win that would boost their bargaining power in state-federal disputes.

–GRETCHEN VOGEL

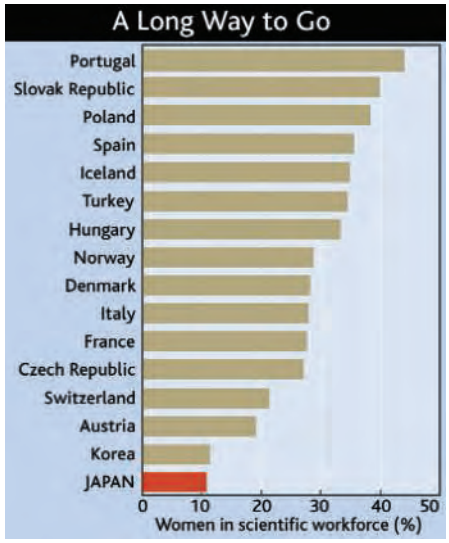
GENDER EQUITY

Japan Mulls Workforce Goals for Women

TOKYO—A government advisory committee has suggested that Japan's publicly supported universities and labs set targets for hiring more women and that the government monitor their progress and publicize the results. The idea is to encourage—and perhaps even embarrass—authorities into lifting Japan from last place among industrialized nations in the employment of women scientists. "We need something to encourage more progress in this area," says Yasuharu Suematsu, former director general of the National Institute of Informatics and head of the panel, which reported this month to the Ministry of Education.

Current figures from Japan's Statistics Bureau show that women make up just 11.6% of the country's R&D workforce. That percentage is the lowest among the 30 industrialized countries in the Organisation for Economic Co-operation and Development (OECD), in which Portugal leads the way with more than 40%. The U.S. figure is 26%.

The advisory committee concludes that raising Japan's percentage will require progress on many fronts, including better support for women with families, gender-blind evaluations, and more aggressive efforts to promote women into leadership positions.



Little women. Japan ranks last in the OECD on women in its scientific workforce.

But as a start, the committee suggests adding targets to the country's next Five-Year Basic Plan for Science and Technology that will govern spending and policy decisions for the half-decade starting next April.

Mariko Kato, an astronomer at Keio University in Tokyo, worries, however, that the targets will lead administrators to boost numbers by hiring "nonassertive women" for low-ranking positions instead of tackling more fundamental problems. "There is still sexual harassment, and you still hear comments about women being unsuitable for science," Kato says. "If you don't change the consciousness of men, the environment for women won't change." Chikako Shingyoji, a female cell biologist at the University of Tokyo who serves on Suematsu's committee, doesn't believe targets are the entire answer. But "setting targets is better than not doing anything," she says.

Suematsu agrees that male attitudes are a big obstacle. "Striving to meet targets will mean addressing the question of how to change this consciousness," he says.

–DENNIS NORMILE

CREDITS (TOP TO BOTTOM): UNIVERSITÄT HEIDELBERG; OECD

Published by AAAS

NASA's new leader comes aboard with a deceptively big budget for science. But size is no substitute for a healthy portfolio. Can Michael Griffin remake the agency's human space flight program without eviscerating research?

Balancing the Right Stuff

Solar physicist Yohei Yamauchi dreams of finding a permanent job in his field. But his boss at the New Jersey Institute of Technology in Newark recently told him that NASA was cutting the modest grant supporting his work analyzing data on the solar corona, leaving the 38-year-old Japanese-born researcher scrambling for another position. A scientist at another research institute who would like to hire Yamauchi is instead laying off a postdoc because of the same budget constraints.

Yamauchi's straitened circumstances are a sign of a quiet crisis in NASA's science program that poses a formidable challenge to Michael Griffin, who took over last week as NASA's new administrator. Space agency managers are now chopping more than $400 million out of the 2005 science budget to cover congressional earmarks and shuttle overruns. That means cutting grants, turning off satellites, and postponing nearly a score of planned missions. And the situation is likely to grow more dire in the coming year, as shuttle costs continue to rise and NASA pushes ahead on programs designed to send humans to the moon and eventually to Mars—all on a budget slated to remain nearly flat.

"There is the potential for serious damage to the future of science at NASA," says Lennard Fisk, a geophysicist at the University of Michigan, Ann Arbor, who chairs the National Academies' Space Studies Board. Fisk, who led the agency's science program during the Administration of President George W. Bush's father, was one of 17 prominent scientists to sign an unusual manifesto the day before Griffin's Senate confirmation hearing urging NASA to retain its broad-based science program while it pursues the human exploration of the moon and Mars. "The balance between the two modes of exploration, human and robotic, is now threatened," the manifesto states.

Griffin—who spoke with some of the concerned researchers a few days before that hearing—echoed that concern at the 12 April hearing. "We as a nation can clearly afford well-executed, vigorous programs in both robotic and human space explo-

"We Can Do the Program That the President Has Proposed"

Calling him "a rare combination of scientist, engineer, and manager," Senator Barbara Mikulski (D–MD) gave voice to the thoughts of colleagues on both sides of the aisle in speeding Michael Griffin through a Senate confirmation process that took all of 1 day. "He is a rocket scientist—thank god we'll have someone who understands what it is all about!" she proclaimed about the new NASA administrator during his hearing on 12 April.

That understanding will be put to the test as the 55-year-old aerospace engineer faces a slew of tough decisions (see main text). Sources close to Griffin predict sweeping changes by this summer in NASA's senior management, including new chiefs of science, space flight, and legislative and public affairs. Their boss has already received White House approval to send a shuttle mission to service the Hubble Space Telescope if he deems it to be safe, they add. During the hearing, Griffin laid out his views on several pressing issues facing the agency.

Here are excerpts from his testimony:

• **On the space station:** "A human space-flight program focused only upon the completion of the space station and the servicing of that station with the shuttle does not qualify as a goal which is worth the expense, the risk, and the difficulty of human space flight. ... The president is pledged and I ... am pledged to bring the space station to a level of completion consistent with our obligations to our international partners."

• **On balancing human and robotic programs:** "If we continue to receive the president's budget allocations, we can do the program that the president has proposed. We know that we can do it because we've done it. The Apollo years are often looked at as a period when the agency had a single mission focus. That [is] incorrect. During the Apollo years, in addition to executing that program, ... we also executed a host of planetary missions in the Mariner, Ranger, Surveyor, Voyager, and Viking Series. We executed earth science missions. ... We executed astronomy missions [and an] orbiting solar observatory."

• **On the Hubble Space Telescope:** "I would like to take the robotic mission off the plate. ... And so I believe that the choice comes down to reinstating a shuttle servicing mission or possibly a very simple robotic deorbiting mission. The decision not to execute the planned shuttle servicing mission was made in the immediate aftermath of the loss of Columbia. When we return to flight, it will be with essentially a new vehicle, which will have a new risk analysis associated with it and so forth. At that time, I think we should reassess the earlier decision."

• **On a new human launcher:** "Two nations [China and Russia] have now put people into space since the United States has last done so. I don't like that. The program that NASA has outlined so far features a new crew exploration vehicle—we can call it what we will—and it nominally comes online in 2014. I think that's too far out. President Bush said not later than 2014. He didn't say we couldn't be smart and do it early. And that would be my goal." **–A.L.**

CREDITS (TOP TO BOTTOM): NASA/HUBBLE HERITAGE TEAM (AURA/STSCI); MARTY KATZ

Published by AAAS

ration as well as in aeronautics," he said. He noted that NASA during the 1960s was not solely focused on the Apollo moon program but had vibrant planetary, earth science, and aeronautics efforts. "We can do it again," he insisted. The next day the Senate confirmed him as the agency's 11th chief.

But Griffin also said his top priorities are getting the shuttle back into orbit and building a new human launcher to replace it. He must contend with major aerospace companies, NASA centers, and key lawmakers committed to preserving the jobs that the space shuttle and space station provide, and a president who wants NASA to push ahead with new launchers, lunar bases, and human missions to Mars. And Griffin, unlike his predecessors in the 1960s, almost certainly will not receive budgets large enough to accommodate these competing interests. "He's going to have to choose sides; he can't make everyone happy," predicts one former NASA administrator. Adds a longtime congressional aide: "He has got quite a challenge to figure out how to make the math work."

Best and worst of times

Ironically, NASA's science program has never been better funded. One-third of the agency's budget—$5.5 billion—is devoted to science. That's the largest percentage in agency history, notwithstanding new accounting methods that include overhead. Construction of sophisticated robots to examine Mars is under way, a large new space-based observatory to replace the Hubble Space Telescope is well along in the planning, a probe to Pluto will soon be launched, and a fleet of Earth-observing spacecraft is returning unprecedented quantities of data. A new lunar robotic effort is on the books, and science's share of the NASA pie is slated to hit 38% in 5 years.

That is little solace to researchers such as Yamauchi, however, who are bracing for more bad news. NASA will soon announce a $160 million cut to its 2005 science budget, after making a similar reduction in December. Another $86 million goes to working on a robotic mission to the Hubble Space Telescope, for a total of $407 million. Later this month, an independent group of scientists will tell NASA which earth science missions should be shut down in light of the funding crunch. And this fall, another panel will determine which half-dozen or more of 13 orbiting solar and space physics spacecraft—including the famed Voyager probes—should be turned off. That advice follows NASA's decision to postpone indefinitely work on most long-term missions that aren't heading to Mars or the moon.

NASA science chief Al Diaz blames the squeeze on congressional decisions, called earmarks, to fund projects not requested by the Administration. "Every 2 years, these earmarks [divert enough money to] eat a mission," Diaz told a NASA earth science and space science advisory committee science advisory panel on 31 March. "Earmark money clearly could have been used to fund Voyager and Ulysses"—two spacecraft currently on the chopping block. Less than a decade ago, such set-asides accounted for just a few million dollars in the science budget.

But many scientists say a far bigger threat to broad-based science at NASA is the rising cost of returning the shuttle to orbit and building the space station, coupled with the president's call last year for human visits to the moon and Mars. "It is only going to get worse," says Princeton University astronomer John Bahcall about raids on NASA's science budget to accommodate human flight. "They will have to dig even more deeply in the science budget; it has only just begun to be mined."

The roots of today's woes were put down soon after the Columbia accident in February 2003, when NASA began the long and expensive job of fixing the shuttle. Meanwhile, the White House developed a long-term strategy for the agency that would finish the space station, shut down the shuttle, and send humans to the moon and Mars using a new launcher that would be ready by 2014. Bush said he would pay for the initiative by phasing out the shuttle in 2010 and abandoning the station several years earlier than originally planned. Although the science focus of the exploration effort would be the moon, Mars, and life science research aboard the space station, then–NASA chief Sean O'Keefe insisted that the overall science program would be protected.

But the Administration's "pay-as-you-go" strategy for its exploration effort, accompanied by modest budget increases for the coming years, began to unravel quickly last year. Although Congress approved the full amount requested by the White House, the costs for getting the shuttle flying again continued to climb, to more than $700 million in 2005 alone, an amount not reflected in Bush's original 2005 request. And Congress packed the NASA budget with pork, including $160 million in the science directorate alone. Meanwhile, cost estimates for robotic missions as well as new technology programs such as the Prometheus nuclear-power system were on the rise.

Confronted with an expensive war in Iraq and a swelling budget deficit, the White House asked for less money in 2006 than Bush had pledged to request just 1 year before. And many aerospace companies and lawmakers object to the president's plan to shut down the shuttle in 2010 when a new human launcher would not be ready until 2014. They argue that the 4-year gap is too long. In his confirmation hearing, Griffin

CREDITS (TOP TO BOTTOM): SPECTRUM ASTRO; NASA; NASA/JPL; NASA/JPL

Published by AAAS

pledged to try to speed up construction of that new human launcher, which would undoubtedly cost tens of billions of dollars. Griffin has previously proposed converting the shuttle from a human launcher into a cargo vehicle, which could also entail a major investment.

Yet neither earmarks nor the human space flight program fully accounts for NASA's science crisis. The Columbia accident occurred as the research community was selling NASA on a new generation of planetary, astrophysics, and earth science missions. To pay for those new programs, the agency planned to spend $1 billion more on science in 2006 than has been requested by the White House.

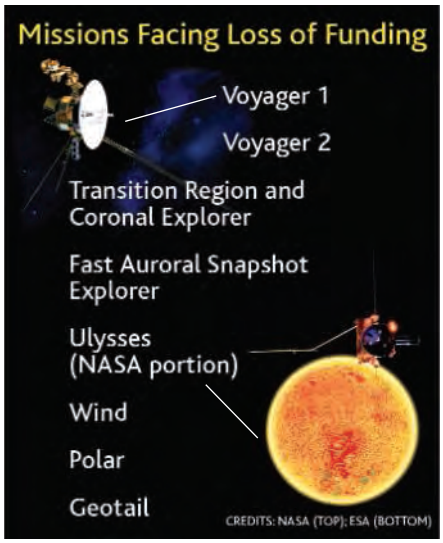
Fisk maintains that those achievements, rather than the president's exploration vision, are largely to blame for the current mess. Diaz's predecessor, Ed Weiler, "was too successful," says Fisk. "He sold programs that required a growth in funding for science that is not now attainable."

Crossing the Rubicon

NASA officials refuse to say exactly how they will allocate the final round of 2005 cuts—a total of $160 million—but the impact is already being felt at U.S. institutes and universities. "I personally elected not to cut ongoing programs and not introduce delays or eliminate strategic programs," Diaz told the NASA advisory panel. "It is impossible to figure out a more surgical way; the problem is our flexibility is gone. Missions have grown in size and funding has not." The chief victims, he said, will be operations of existing spacecraft, grant programs, and longer-term plans to build earth science and astrophysics probes.

In NASA's astrophysics division, for example, managers are struggling to cope with costly technical troubles on spacecraft slated for launch in the next few years along with cuts imposed from above. Anne Kinney, the division leader, late last year squeezed $100 million from her 2005 budget of roughly $1.5 billion to cover earmarks, general reductions, and returning the shuttle to flight. In the past month she has had to find another $58 million in reductions—a task made harder by the fact that the budget year is already more than half over. Two-thirds of the reductions will be assigned to missions and one-third to research and analysis programs. "As long as I've been here, we've never cut research," says Kinney. "We are crossing the Rubicon."

The impact was immediate. The same day that Diaz spoke to his scientific advisers, NASA announced it would cancel this year's solicitation for 5-year, $100,000 grants that allow budding astrophysicists to pursue a broadly framed scientific issue.



"For young researchers like me, these long-term programs are absolutely vital," says one grantee, Bryan Gaensler, a 31-year-old assistant professor at the Harvard-Smithsonian Astrophysical Observatory in Cambridge, Massachusetts.

Kinney says she had little choice. "Why ask 200 people to send in proposals if you can only accept five?" she adds. NASA is canceling another annual grant program that funds work on data archives from older missions. The agency also put on hold other efforts like the Explorer programs, which fund modest missions from a variety of disciplines at a faster pace than the usual NASA projects. That outrages many space scientists. "This is the lifeblood of innovation and creativity in our discipline," says Fawwaz Ulaby, an electrical engineer at the University of Michigan, Ann Arbor, and member of NASA's new science advisory committee. "Our community has been saying that the Explorer program is absolutely critical, so that we have some agility to respond" to new research questions.

Diaz also intended by this fall to turn off seven of 13 operating solar physics missions—including the Voyagers at the edge of the sun's influence—to save a total of $21 million. The probes represent the bottom half of a 2003 ranking of scientific usefulness. But under pressure from the advisory panel, he recently agreed to conduct an extensive outside peer review this autumn before terminating any missions. That extension, warns NASA manager Paul Hertz, puts further internal pressure on the science budget.

Down to Earth

Earth scientists hope that their work on a National Academies' decadal plan—an interim version was slated for release this week—will help them persuade the White House and Congress not to abandon their troubled discipline. It will be an uphill struggle. Berrien Moore, a biogeochemist at the University of New Hampshire in Durham who is a co-chair of the decadal panel, calls NASA's current approach "a going-out-of-business sale for earth sciences."

He notes that only one of a half-dozen missions planned for launch is clearly going forward. NASA, for example, plans to abandon the Glory mission to study aerosols, a mission championed just last year by former NASA chief Sean O'Keefe, who pledged to speed up launch to 2007 or 2008. "Now we have gone from acceleration to cancellation," Moore adds. Some planned missions are indefinitely postponed; others were left with money to build an instrument but without funding for a spacecraft to fly on.

Nor are existing satellites safe. Mary Cleave, who heads the earth science division, predicts that an independent review of several missions to be finished this month will lead to the termination of some of them. At

CREDITS (TOP, LEFT TO RIGHT): NASA/JPL; NASA; NASA; L-3 COMMUNICATIONS INTEGRATED SYSTEMS

Published by AAAS

the 31 March advisory meeting, Harvard University atmospheric chemist Daniel Jacob warned that the earth science program is being "decimated." Diaz called that criticism "a little bit of an overstatement." But he acknowledged that NASA was focused on "strategic issues," shorthand for an emphasis on lunar and Mars exploration along with space station life sciences.

Diaz's policies haven't gone over very well in the space and earth sciences communities. "This is probably just not good budget strategy," says Fisk. Faced with similar constraints in the early 1990s, Fisk chose to scale back efforts to build large spacecraft to protect more fragile smaller missions, existing spacecraft, and the network of scientists who depend on NASA grants to analyze data. "You can't just fund the flight programs, which mostly funnel money to industry," he adds. Researchers have yet to grasp the severity of the issue, and Fisk is worried about their reaction. "There is a firestorm coming, and the community does not always respond in an organized way," says Fisk.

The first organized response is the 1500-word manifesto timed for Griffin's confirmation hearing. It is the brainchild of Nathan Schwadron, a 36-year-old space physicist at the Southwest Research Institute in San Antonio, Texas, who says he began to worry last fall that NASA's science program—and entire disciplines—were in jeopardy.

The document argues that Bush's exploration vision shouldn't be confined to the moon or Mars. "Should other forms of space exploration be canceled or curtailed to make this new, but limited, exploration vision possible? We think and hope not," says the paper. "It is critical that we continue to explore broadly."

Schwadron and others say that they are sympathetic to a revamped human space flight program, but that they want to ensure a broader definition of exploration. "It's not the turning toward exploration, it is the turning away from science that's the problem," says Yale University astronomer Meg Urry. "Some of the most successful science at NASA is languishing, such as the search for dark energy, arguably the biggest revolution in physics in a century." Scientists like Schwadron and Urry applaud the new goals for the human space flight effort, but they don't want NASA's diverse research portfolio to shoulder the costs.

NASA managers insist that the president's vision is fundamentally friendly to science. "Science activities are built into the foundation of the exploration vision," James Garvin, NASA's chief scientist, told the American Astronautical Society at a 29 March meeting in Greenbelt, Maryland. He argues that exploration "is a scientific journey," citing the ambitious plans for lunar and Mars exploration. And Diaz notes that a series of "road maps" being assembled will lay out the long-term direction of science and be completed in time to influence the 2007 budget submission this fall.

As he settles into his ninth-floor offices overlooking the Potomac River in downtown Washington, D.C., Griffin must decide how to balance the fiercely competing needs of the traditional space-flight program, the president's new vision, and science involving more than the moon or Mars. The agency's present course, Schwadron predicts, could eventually force a third or more of the people in solar and space physics out of the field. Astrophysicists, biologists, astronomers, and earth scientists express similar concerns. And younger researchers eager for a stable future are getting skittish.

Scott MacIntosh, a solar physics postdoc at the Southwest Research Institute, can't ignore rumors that the guest investigator program that funds his work may disappear. "I have a background in medical imaging, so I might try to do more cross-disciplinary work," he says. And MacIntosh is in no position to gamble about his future: "I've got a young kid and another on the way." NASA's ability to cultivate a new and diverse generation of space scientists like MacIntosh and Yamauchi may hinge on whether Griffin has the right stuff to execute a difficult balancing act.

–ANDREW LAWLER

Ecosystem Management

California Tries to Connect Its Scattered Marine Reserves

Researchers hope that new funds, better management plans, and the latest science will help them establish the largest network of marine protected areas in the U.S.

The 1800-kilometer California coastline supports a spectacular diversity of marine life. So does a sea floor that plunges just offshore to nearly 2600 meters, with sea-grass beds and kelp forests giving way to submarine canyons and deep rock reefs. Add in seasonal winds and complex ocean currents that churn up nutrients for thousands of species from sharks and tuna to squid and rockfish, and the result is an incredibly rich ecosystem—and one of the most productive fisheries in the nation. Can the two coexist? Those working on a new state effort to create a network of marine protected areas (MPAs) hope that the answer is yes. But it won't be easy.

Once upon a time, the bounty of the sea accommodated both fishers and conservationists. But over the past 2 decades, fish catches have fallen by more than half. An MPA network would set aside part of the ocean to prevent the total degradation of this habitat, foster marine diversity—and perhaps maintain a sustainable fishing industry. The vision is grand. Not only would the network be the largest such system in the nation, but its success "would be a wonderful model" for a national system, says Jane Lubchenco, a marine ecologist at Oregon State University in Corvallis. And last week the group reached its first major agreement: choosing the location of a pilot project.

Location, location

California set up its first MPA in 1957, a 35-hectare area near La Jolla in San Diego county. Since then, 104 areas have been added in a piecemeal and uncoordinated fashion. Despite this effort, MPAs cover less than 0.3% of state waters—not enough to make a difference in helping fisheries recover, scientists say. And none of the reserves protects species or habitats in deeper water.

In 1999, California tried to address the problem with the Marine Life Protection Act

Hooked? Fishers are worried about the economic impact of new reserves but are playing along for now.

CREDIT: RICHARD ALLEN

Published by AAAS

(MLPA). One of the law's requirements is to create a network of MPAs along the state's coastline. But a 2002 attempt by the state department of fish and game and outside scientists encountered stiff resistance from commercial and recreational fishers, who objected to the size and area of the proposed MPAs and the fact that they had been largely excluded from the process. One year later an expanded task force tried again, but it ran out of funding before finishing its work.

Last year state officials tapped into a group of foundations led by the Resources Legacy Fund Foundation (RLFF). RLFF is providing most of the $9 million of funding for the new effort—enough to get a first collection of MPAs set up. Its 19-member science advisory team—appointed by an MLPA task force—is chaired by Stephen Barrager of Stanford University and includes fishers and others.

The group's first challenge was to outline the steps to pick locations for reserves, design their boundaries, and specify how they should be monitored. The guiding principle was that reserves will be more effective as a network rather than isolated, an idea espoused by a 2001 report by the National Academy of Sciences. A network takes into account the movements of adult and larval fish, allowing fish to travel from one reserve to another.

The California plan calls for locating MPAs based on how fish species migrate and how far their larvae disperse. Scientists are learning those patterns in several ways, such as by tracking larvae and analyzing their DNA from various locations. Other experts such as Mark Carr, a marine biologist at the University of California, Santa Cruz, are conducting tagging studies and analyzing fish otoliths to track the movements of adult fish. His data suggest that the ranges of many rocky reef fishes are less than 5 km. But other fish are known to swim about 10 km a day—implying that the reserves ought to extend into federal waters.

The plan breaks new ground with its emphasis on deeper-water habitat. It would protect five types of habitat—including submarine canyons and deep rock reefs—at four depth zones. "Many fish use kelp forests as nursery, move deeper as they grow, and return to shallow water for spawning," says Carr.

The ultimate shape of the MPA will be determined using bathymetric data, maps of nutrient upwellings, and information on the variety and abundance of species in a habitat. Although previous reserves were irregularly shaped, the plan calls for new ones to follow lines of latitude and longitude—making it easier for fishers to avoid them and enforcement officers to scout for violators. There's no guarantee of sufficient funding for officers and equipment, however. As for monitoring, state fish and game scientists should evaluate MPAs by regularly checking species abundance, habitat quality, and other biological indicators.

Teeming. Reserves will be sited where winds dredge up nutrient-rich water (*inset*; blue and violet) that sustains gopher rockfish (*above*) and other species.

Last week, the task force chose a 300-km stretch of the central coast region, roughly from northern Monterey Bay to Santa Barbara, as its first region of study. Existing MPAs within the boundary will be evaluated and weighed in conjunction with any proposed new sites. The first network could be operational by March 2006.

Rough seas

Although participants say the process has been smooth sailing to this point, they expect the political seas to become more choppy when it comes time to decide the exact location and size of the MPAs. At stake is an estimated $1.4 billion sport and commercial fishing industry, an industry already besieged with dwindling fish stocks, decreasing catches, and increasing regulations. Despite being included, commercial and recreational fishers still worry that large chunks of the ocean could be marked off-limits, says Thomas Raftican, president of the United Anglers of Southern California in Huntington Beach. Indeed, Carr's analysis suggests the need for reserves significantly wider than the 1 km that is now typical.

One of the most politically contentious and unresolved problems is striking a balance between areas where fishers have quotas and those from which they are banned. Although protecting 20% has become a commonly cited target, some experts suggest that more than 35% of the areas should be no-take zones, citing successful practices in the Great Barrier Reef. Carr hopes to defuse the issue by noting that the size and shape of each MPA will depend on the species, the habitat, and the conservation objectives of both the individual MPAs and the entire network.

Another major worry is whether the MPAs will boost stocks outside the reserve sufficiently to benefit fishers and win their support. Lubchenco says this "spillover" is likely in California, pointing to increased fish catches outside the Great Barrier Reef and the Florida Keys National Marine Sanctuary. According to Steven Berkeley, a research biologist with the Long Marine Lab at the University of California, Santa Cruz, reserves lead to older and fatter fish, which sustain fish populations by producing more hardy larvae. Early results from a network of 12 MPAs set up in 2003 in the Channel Islands suggest that fish populations outside the reserve are indeed on the rise. But others are skeptical. "There simply are no benefits to commercial fisheries," contends Raymond Hilborn, a fisheries management expert at the University of Washington, Seattle.

Despite these uncertainties, almost everybody agrees that MPAs have the potential to be an important tool in marine conservation and fishery management. "They could act as a buffer or insurance against overfishing or a natural disaster," says Peter Sale, a tropical marine ecologist at the University of Windsor, Canada. People also agree that plowing full steam ahead is the only logical next step. **–AMITABH AVASTHI**

CREDITS: (INSET) NOAA COASTWATCH PROGRAM; MARK CARR

Published by AAAS

Crisis of Confidence Hampers Marburg Control in Angola

Experts have everything they need to stop the deadly Marburg outbreak in northern Angola—except trust from the local population

Vincent Brown has his own way of keeping track of the Marburg virus in Uige, a provincial capital in northern Angola: He counts fresh graves. A daily visit to the town's cemeteries doesn't yield precise numbers, says Brown, an epidemiologist with Epicentre, the Paris-based research arm of Médecins sans Frontières (MSF)—but it does give one a feeling for the trend.

The reason behind the unorthodox method is simple. In the current outbreak of Marburg hemorrhagic fever, which had caused at least 227 deaths by 15 April, most patients never make it to the hospital. Widespread fear and mistrust of public health authorities and the international teams fighting the disease are leading most families to keep their patients at home. As a result, the virus keeps festering, says Brown, who returned to Paris last week from Uige.

Four weeks after Marburg was nailed as the culprit, the fight against the virus has become a battle to win the trust of the local population. "It's clearly a bit more difficult than we anticipated," says Pierre Rollin of the U.S. Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, which has sent several teams to Angola. But medical anthropologist Barry Hewlett of Washington State University in Vancouver says the difficulties were predictable. "It's often like this," says Hewlett, who has accompanied medical teams during several outbreaks of Marburg's cousin, Ebola.

Marburg, which spreads through direct contact with blood and other bodily fluids, isn't like flu, measles, or other highly contagious viruses. Putting patients in strict isolation and checking their close contacts for symptoms daily for at least 21 days—and isolating them as well if they do get sick—will usually end the transmission chain.

Today, the logistical systems are in place to do just that, says Pedro Pablo Palma of MSF's Spanish branch in Barcelona. MSF has set up a three-compartment isolation unit in the hospital in Uige, for suspected, probable, and confirmed patients. Although hygiene in the rest of the hospital was initially "catastrophic," Pablo says, with highly infectious corpses piling up in the morgue, the situation has gradually improved.

Scientific capacity is generally better than in most previous outbreaks of Marburg and Ebola. Researchers from Canada's National Microbiology Laboratory in Winnipeg have set up a field lab in Uige that can test patient samples within a few hours. CDC, meanwhile, has set up a diagnostic lab in Angola's capital Luanda to test any samples that might come in there and to confirm results by the Canadian team. But while the graveyard kept filling up—there were twice as many new graves daily in early April than in March, Brown says—the labs didn't have nearly as many samples as they could have handled, and the isolation unit was virtually empty early this week.

The lack of trust has several roots. One is that so far, nobody has made it out of the isolation unit alive, says Pablo—not surprising with a fatality rate of close to 100%, at least of those who make it to the hospital. The notion of isolation itself has been hard to accept, adds David Daigle, a CDC communications officer acting as a spokesperson for the World Health Organization (WHO) in Angola. And at the outset, deceased patients were immediately zipped into plastic bags to prevent further infections, Daigle says, even though tradition requires a ritual washing of the body, during which the deceased is embraced or kissed. "People were very upset," he says. "They couldn't grieve."

The result has been not just a lack of cooperation but also outright hostility—not so much in the city but in four or five of its 14 suburbs, says Brown, who was chased away by an angry mob of 40 to 50 people after a visit to a traditional chief, or soba, in one of them. "It felt pretty threatening," he says. "The message was: Don't come back here."

For now, WHO and MSF are heeding that message and shunning certain areas in the hope that a broad "social mobilization" campaign will soon change attitudes. To that end, sobas, church leaders, and traditional healers are being recruited. Two medical anthropologists—one from France, the other from Burundi—are helping with this process, says Daigle.

Some creativity is clearly needed. To replace the traditional washing ritual, the anthropologists have introduced an alternative in which family members sprinkle the dead body with bleach, says Daigle. And a popular band whose lead singer died from Marburg has recorded a song to help raise awareness; trucks mounted with loudspeakers should be blaring it out soon.

If past experience is any guide, such measures can usually win over a population, as long as they are culturally sensitive and build on existing beliefs, says Hewlett. In recent Ebola outbreaks in Uganda and the Republic of the Congo, certain changes in burial rituals were generally accepted, such as wearing plastic gloves or introducing bleach. Simply putting bodies in plastic bags was a big mistake, Hewlett says, however well intended.

Still, he's not surprised. Sometimes, the teams sent out to hemorrhagic fever outbreaks are a bit like "medical cowboys," he says. "They feel very strongly about what they have to offer, especially in a crisis"—and fail to realize it may not always be appreciated.

—MARTIN ENSERINK

Staying away. Marburg patients aren't coming to an isolation unit in Uige.

CREDIT: FLORENCE PANOUSSIAN/AFP/GETTY IMAGES

Published by AAAS

MILWAUKEE, WISCONSIN—About 1200 researchers gathered near the shores of Lake Michigan here from 5 to 9 April to discuss early Englishmen, the birth of modern humans, and Stone Age weapons.

Once-Balmy Climate Lured Humans to England Early

Scientists following a trail of stone tools and butchered animal bones have uncovered evidence that early humans lived in Britain well before 500,000 years ago, perhaps not long after the first Europeans appear much farther south in Spain and Italy, about 800,000 to 1 million years ago. The early English settlers probably followed a wave of hippos, elephants, hyenas, and other animals drawn to Britain's then-balmy climate, according to a talk and poster by paleoanthropologist Chris Stringer of the Natural History Museum in London. But when the climate cooled, as it did repeatedly over the following epochs, all traces of human occupation vanished.

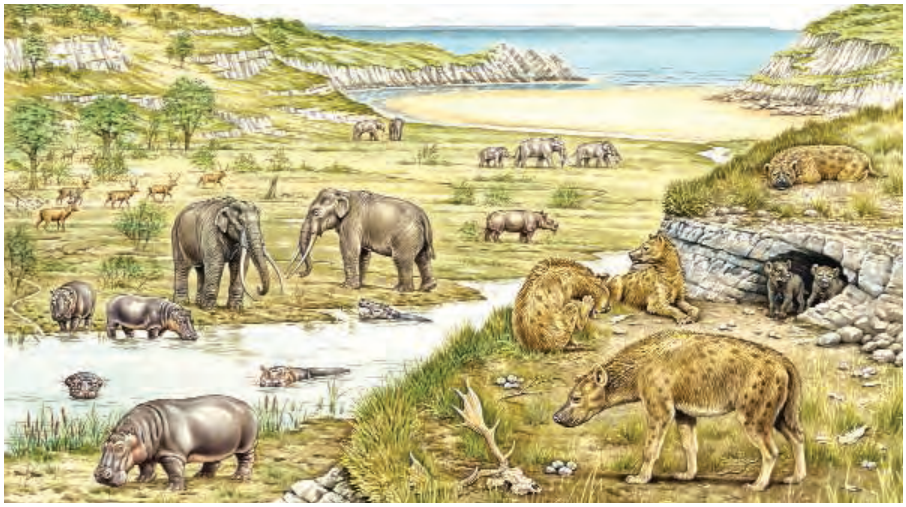
Several new sites suggest that humans were in Britain well before the appearance of the 500,000-year-old Boxgrove Man, whose shinbone and teeth were discovered in a gravel quarry in Boxgrove, England, from 1993 to 1996. The sites may help shed light on whether more than one type of human migrated to Europe more than 500,000 years ago and reveal the type of terrain they could inhabit. "This pushes the age of humans north of the Alps back further than previously documented," says paleoanthropologist Erik Trinkaus of Washington University in St. Louis, Missouri.

Boxgrove showed that the earliest known Briton was a member of *Homo heidelbergensis*, a proto-Neandertal species with deep roots in Europe. The new sites have no human remains, but researchers found tools along the coast of the ancient Bytham River in East Anglia. The tools appear in some of the most ancient river terraces and are associated with insects and animals that suggest a date far older than Boxgrove, Stringer said in his talk. One site with tools may be as old as 700,000 years.

These early Europeans carried a primitive stone tool kit for scraping and cutting. But they lacked the hand ax—a versatile stone tool nicknamed the Paleolithic Swiss Army knife—already in widespread use in Africa. The Boxgrove hominid did wield a hand ax and so may have been part of a separate wave of settlers, says Stringer, who directs the $1.88 million Ancient Human Occupation of Britain program funded by the Leverhulme Trust. Studies of animal fossils paint a portrait of a warm climate that allowed animals now found only in Africa to migrate from northern Europe to England across a land bridge.

Although humans arrived in Britain early, they did not live there continuously, said Stringer. There are no signs of human occupation during several periods, particularly during glaciations. From 180,000 to 130,000 years ago, herds of mammoth and reindeer roamed England, but there is little evidence of humans. Hippos and elephants reappear when the ice caps melt at about 130,000 years, but humans don't show up again until about 60,000 years ago when Neandertals return. Modern humans came later, but even they disappeared during an Ice Age as recent as 25,000 to 17,000 years ago. "People assume that once people were in Britain, they were always there," says Stringer. "We're seeing little pulses of human occupation. They disappeared when it got very cold. There is not a continuous human presence until 12,000 years ago."

–ANN GIBBONS



CREDIT: JOHN SNOWBALL

English summer. Humans trailed mammals such as hyenas into England more than 500,000 years ago.

Archaic Genes in Modern People?

In the past 15 years, a flood of genetic data has helped propel the Out of Africa theory into the leading explanation of modern human origins. DNA from mitochondria (mtDNA), the Y chromosome, and ancient humans each suggest that the ancestors of all living people arose in Africa some time after 200,000 years ago, swept out of their homeland, and replaced archaic humans around the globe without mixing with them. But at a genetics symposium, two independent groups presented data from the X chromosome hinting that modern humans interbred with other human species: The teams found possible traces of archaic hominids in our genes. "Just as the Y and mtDNA data seemed to have settled it, the new data revive the question [of interbreeding]," says Stanford University's Joanna Mountain, co-organizer of the symposium. "The controversy is not settled."

Geneticists Makoto Shimada and Jody Hey of Rutgers University in Piscataway, New Jersey, presented an intriguing haplotype—a set of genetic mutations inherited together—that appears to have ancient roots in Asia rather than Africa. Shimada sequenced a 10.1-kilobase noncoding region in 659 individuals from around the world. Overall, the genetic variations were most frequent in Africa, just as expected if our ancestors were a subset of ancient Africans who migrated out of that continent. But one rare variant, appropriately named haplotype X, appeared in nine individuals from Europe to Oceania but was entirely absent in Africa. Shimada estimated that the haplotype arose 1 million years ago, long before the modern human exodus from Africa. "Haplotype X is difficult to explain by the recent African origins model," says Shimada. "It's very old, it's rare, and it is widespread outside of Africa."

In independent work, geneticist Michael Hammer of the University of Arizona in Tucson offered a similar example. Hammer and postdoc Dan Garrigan identified a 2-million-year-old haplotype in the RRM2P4 region of the X chromosome that is common in East Asia but vanishingly rare in Africa. Their work, published 2 months ago in *Molecular Biology and Evolution*, raises the possibility that the haplotype arose in very ancient Asian popula-

Published by AAAS

tions, presumably of *Homo erectus*, an ancient human once found across Asia. "This is what you'd expect if you had introgression" between modern humans and *H. erectus*, Hammer said.

But at this point several other explanations are possible. Hey of Rutgers acknowledges, for example, that haplotype X may be present in Africa but was missed by spotty sampling in that continent. "Simply observing those [examples] is not sufficient to rule out one model or another," cautions Mountain. "What you need is 10 or 50 loci—one or two is not sufficient." Hammer, for one, thinks that these preliminary data do "speak to some archaic admixture. The few [loci] we've done so far are so suggestive that it gives me great excitement to continue sequencing more loci."

–ELIZABETH CULOTTA

Snapshots From the Meeting

Speed demon. Once thought slow, the 130-gram red loris was filmed darting through shrubs.

New view of lorises. The tiny, nocturnal lorises have been considered the sloths of the primate world, creeping carefully along the shrubbery of their rainforest homes. They've also been considered a skinny branch of the primate tree, with fewer than 10 species described. But at the meeting, lorises emerged as surprisingly speedy and speciose. Anna Nekaris of Oxford Brookes University in Oxford, U.K., showed a field video of the endangered red loris scrambling around Sri Lankan trees at about 1.3 meters per second, twice as fast as captive animals. Other presenters argued that researchers have missed variation in the hard-to-track lorises: Subspecies vary in size by as much as 50%, with many differences in craniofacial proportions, says Matt Ravosa of Northwestern University Medical School in Chicago, Illinois. Jeff Schwartz of the University of Pittsburgh in Pennsylvania adds that "subspecies" differ in skull and tooth shape, too, and predicts that some will be identified as species soon.

Human relations. Sarah Tishkoff and Floyd Reed of the University of Maryland, College Park, presented preliminary analyses of a massive data set on genetic variation in humans around the world, particularly Africans. Samples from more than 3000 people, including 2000 Africans, were processed at 1275 loci by a genotyping powerhouse, the Marshfield Clinic Research Foundation in Wisconsin. Tishkoff and Reed, who received the complete data set only 3 weeks ago, say it offers a powerful tool to uncover relationships among populations. For example, the data suggest that culturally distinct groups of Pygmies are more closely related to each other than to other Africans. The researchers also detected unique similarities in the peoples of Oceania and East Africa, lending support to the hypothesis of an early "southern route" of migration out of Africa, around the coast of India to Oceania and then Australia. Finally, they found ancient kinship among three groups of click speakers, supporting the idea that the click languages form a single, ancient language family (*Science*, 27 February 2004, p. 1319).

–E.C.

Modern Humans Made Their Point

Long before guns gave European explorers a decisive advantage over indigenous peoples, our ancestors had their own technological innovation that allowed them to dominate the Stone Age competition: the projectile point, launched from bows or spear throwers. Paleolithic hunters shooting spears or arrows tipped with these small stone points could stay at a safe distance while hunting a wide assortment of prey—or other humans, says archaeologist John Shea of Stony Brook University in New York. Projectile launchers might even be the key to modern humans' triumph when they entered the Neandertal territory of Europe about 40,000 years ago, Shea proposed in his talk. Neandertals lacked projectiles until it was too late, and they could heft their heavier spears only as far as they could throw them. "Projectile points were such an important invention, like gunpowder, that it would have given the bearers a huge advantage," says archaeologist Alison Brooks of George Washington University in Washington, D.C.

In two separate studies, Shea and Brooks showed that modern humans were using lightweight points associated with projectile launchers by 40,000 years ago. Shea and Brooks both think these new weapons were invented first in Africa, although they disagree about the timing. They agree that modern humans had a technological advantage when they left Africa and spread around the globe. "These lightweight points show up more than 50,000 years ago in Africa," says Stan Ambrose of the University of Illinois, Urbana-Champaign, who heard Shea's talk. "They may have helped modern humans get out of Africa."

The challenge in pinpointing when projectiles were invented is that few of the launchers themselves survive, because they were made of materials that disintegrate over time. The oldest known bow is only 11,000 years old, and the oldest knownspear thrower is about 18,000 years old, but archaeologists suspect that the technology is much older. So they try to distinguish projectile points from those used on the tips of hand-thrown spears. One criterion is size: Projectile points must be small and light to soar fast enough to kill. "You wouldn't go up to a Cape buffalo with those tiny points on a thrusting spear," says Brooks.

The modern edge. Launchers shot arrows tipped with small blades (*center and right*).

Shea and Brooks each surveyed points from around the world, setting an upper limit on the size and weight of points considered projectiles. Shea set an upper limit on cross sections at the tip, whereas Brooks set a limit on weight. Shea found that projectile points were widespread by 40,000 years ago; earlier points didn't meet his criteria. He proposed that the points were developed for warfare and may have hastened the extinction of Neandertals.

Brooks found that points from 50,000 to 90,000 years ago in three regions of Africa met her criteria. She noted that there was a "grammar and an order" to assembling these tools—one that required extensive social networks in order to exchange technology and specialized materials. She thinks that projectiles made modern humans more efficient hunters who could shoot small game and live in varied terrain. "They didn't have to kill [Neandertals]," says Brooks. "They just had to outcompete them."

–ANN GIBBONS

CREDITS (TOP TO BOTTOM): K. A. I. NEKARIS/OXFORD BROOKES UNIVERSITY; JOHN SHEA/STONY BROOK UNIVERSITY

Published by AAAS

Random Samples

Edited by Constance Holden

Old Coot

Perfectly preserved in silica, feathers and all, this 3D fossil depicts an American coot that met its death in one of Yellowstone National Park's hot springs between 5000 and 10,000 years ago. The bird, discovered by taphonomist Alan Channing of the University of Cardiff, Wales, and colleagues, is the first avian fossil to be found in a hot spring and one of the few vertebrates.

Such fossils are rare, says Channing, because "soft tissues get destroyed very quickly" by microbes and chemicals in the springs. But in the case of the coot, corpse-colonizing microbes appear to have sped up a process of encrustation from the surrounding silica, leaving a perfect cast of the bird, the researchers report online on 13 April in *Proceedings of the Royal Society, Series B*. Because soft tissue is generally not found in fossils, even "a one-off specimen could really answer some questions" about the lifestyles of ancient birds, says Channing.

Venice Plans Sublagoon Tube

After more than a century of discussion, Venice is on the verge of approving the construction of a subway under its lagoon to ease its canal-bound transportation system. The measure should also help protect the city's ancient buildings, which are set on wooden foundations: Waves created by canal traffic "damage their delicate structure," according to the city council. If plans made public this month get the green light, a "sublagoon transport system" will start operation within 6 years. The $450 million project awaits funding pending local government approval of an environmental impact study.

Dotted line marks proposed track.

The single-line metro is designed to whisk travelers from the mainland-based airport, going underground at the lagoon and passing through the island of Murano before terminating on the northeast side of Venice. The proposal has divided opinion. Some environmentalists worry that the project, which would entail digging through consolidated sediment 20 meters below the lagoon floor, will cause further subsidence in the already-sinking city. But the city's outgoing mayor Paolo Costa says the subway would take tourist pressure off the ferries and "give Venice back to the Venetians."

Colored Memory

When Daniel Tammet set the European record for pi memorization last year, absorbing 22,514 digits in just over 5 hours, he attributed the feat to his ability to see numbers as complex, three-dimensional "landscapes," complete with color, texture, and sometimes even sound.

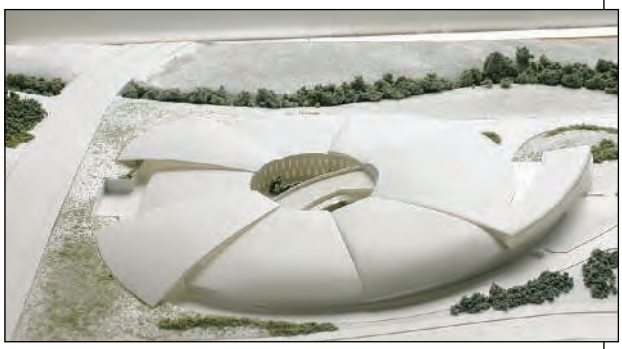
To see whether this form of synesthesia is at the heart of Tammet's talent, neuroscientist Vilayanur Ramachandran and colleagues at the University of California, San Diego, gave the 26-year-old savant from Kent, U.K., a series of tests. He had 3 minutes to memorize 100 digits and their locations in a 10-by-10 array. When the digits were all the same size, Tammet recalled 68 correctly, and he remembered all 68 when tested again 3 days later. But when the test was given again with digits of different sizes to disrupt Tammet's synesthetic imagery, his performance plummeted to 16 correct, and zero 3 days later, according to a poster presented 10 April by Ramachandran's student Shai Azoulai at a meeting of the Cognitive Neuroscience Society in New York City.

The team now plans to investigate the multiplication skills of Tammet, who says he visualizes the shapes of the numbers to be multiplied and then reads off the product from a third shape that appears in the space between them. The researchers want him to produce a set of number shapes, in clay or on a computer, so that they can uncover principles governing his number representation.

"It's an extremely interesting idea" that such vast memory capability can be supported by synesthesia, says Lynn Robertson, a cognitive neuroscientist at the University of California, Berkeley. Little is known about memory tricks used by other savants because they tend to express little insight into their talents, says Ramachandran.

Spanish Synchrotron

The shape of Spain's first synchrotron was unveiled early this month. The winning design in a competition for the €165 million, 3-giga-electron-volt radiation source is a novel snail-like structure that will organically "allow for future expansions," says physicist Joan Bordas, director of the ALBA Synchrotron, as it's called. Construction will start in Barcelona early next year. The facility is supposed to open for business in 2009.



CREDITS (TOP TO BOTTOM): ALAN CHANNING/CARDIFF UNIVERSITY; LANDSAT; ALBA

Published by AAAS

Edited by Yudhijit Bhattacharjee

JOBS

New head at Sandia. Weapons engineer and longtime lab employee Thomas Hunter has been named director of Sandia National Laboratories, the Department of Energy's weapons research facility in Albuquerque, New Mexico. He succeeds C. Paul Robinson, who joined Lockheed Martin this month to bolster the company's bid to also run Sandia's in-state neighbor, Los Alamos National Laboratory (*Science*, 15 April, p. 339).

Hunter, 59, joined Sandia in 1967 and has worked on underground testing, waste management, and energy and environmental programs. He says he'd like to see the lab's expertise in areas such as high-performance computing broadened from weapons to other "defense applications like homeland security."

Hunter is "absolutely passionate about national security," says former Sandia chemist Al Sylwester, who helped Hunter build partnerships between Russian and U.S. weapons laboratories. Hunter takes over next week.

POLITICS

Change of heart. Missouri scientists who favor human embryonic stem cell research found an unlikely political ally this month. In an hourlong floor speech, Christopher Koster, a freshman Republican state senator, cited science and Scripture to help kill a bill that would have outlawed somatic cell nuclear transfer.

"The Psalms tell us, 'He knit me together in my mother's womb,'" Koster, 40, told his colleagues. "The National Institutes of Health tells us a human embryo exists from the time of implantation until the end of the eighth week."

Koster voted in February to move the same bill to the floor, although he claims he was unsure of his stance at the time. Subsequent talks with religious mentors and scientists convinced him that the "human miracle" of normal embryo development was different from research cloning. That decision made him a "hero on the floor" to business leaders, scientists, and patient groups who had fought the bill for months, says lobbyist Rose Windmiller of Washington University in St. Louis.

Headed out. The end is in sight for Philippe Kourilsky, the embattled head of the Pasteur Institute in Paris. A new board of directors has decided that Kourilsky will not be asked to serve a second term and should even step down before his current 6-year mandate ends in December if a replacement can be found.

The immunologist's attempts to rejuvenate Pasteur have been hampered by an authoritarian management style and a controversial plan to partially move the lab to a Paris suburb (*Science*, 4 March, p. 1391). The problems have made him a lame duck at a time when a major renovation and other important decisions are pending, says microbiologist Agnès Labigne, secretary of the board of directors. "We have to find a new director as soon as possible," she says, adding that a search committee might start work this week. A Pasteur spokesperson said Kourilsky was traveling.

Labigne says the new president should avoid micromanaging the institute and prefer fundamental research over ties with industry.

NONPROFIT WORLD

Broader role. Six weeks after accepting a second 5-year term as president of Berlin's Humboldt University, physicist Jürgen Mlynek has agreed instead to become president of the Helmholtz Society, Germany's largest research organization. The society, with a budget of $2.75 billion this year, governs 15 of Germany's largest science institutes, including the German Synchrotron Research Centre (DESY) in Hamburg and the Max Delbrück Center for Molecular Medicine in Berlin.

Mlynek's departure has prompted accusations of disloyalty at Humboldt, where he had promised to continue reforms aimed at pushing the university to international prominence. He says he understands his critics but adds that the new position strengthens his ability to tackle "the needs of research and education in the German universities."

Mlynek says he gave up research when he became president at Humboldt but hopes to retain ties to his former lab. "Whenever I am strongly frustrated, I go over and talk to the students, and I feel better," he says. He takes up his new job in autumn, succeeding Walter Kröll, who is retiring.

Got any tips for this page? E-mail people@aaas.org

THEY SAID IT

"I won the lottery. Most people in my situation would have died, and I got a really lucky break."

—Nobelist Eric Cornell, who returned to work at the National Institute of Standards and Technology in Boulder, Colorado, this month after surviving necrotizing fasciitis. Cornell's left arm and shoulder had to be amputated because of the infection, which is caused by a flesh-killing bacterium.

CREDITS (TOP TO BOTTOM): SOURCE: CHRIS KOSTER; SOURCE: SANDIA; INSTITUT PASTEUR; DAVID AUSSERHOFER/HUMBOLDT UNIVERSITY

Published by AAAS

Evolution Can't Be Taught in 270 Minutes

WE WERE PLEASED AND GRATIFIED TO READ that Jennifer Miller chose to walk out on her ninth-grade biology classes rather than read the unnecessary and misleading statement foisted upon her by the school board ("Dover teachers want no part of intelligent-design statement," J. Mervis, News Focus, 28 Jan., p. 505). However, we were dismayed to read that she "will spend at most three 90-minute classes on the topic—the last unit of the year…" Evolution is not a "unit." It is the greatest unifying theme in all of biology and must be incorporated from day one throughout the academic year.

For six years, we conducted a series of graduate institutes for science teachers entitled "Evolution and the Nature of Scientific Inquiry: Using Evolution as a Central Theme in Life Science Courses." Seven recommendations emerged: (i) Science teachers should be required to take a course in the history and philosophy of science. (ii) Evolution needs to be addressed early in the educational system in a nonapologetic, noncontroversial fashion (*1*). (iii) Undergraduate courses in the life sciences should be taught with an overt evolutionary theme. (iv) Life science teachers should be required to take a course in evolution. (v) Life science textbooks need to be written with the permeating themes of the nature of science and evolution. (vi) Science teachers must cover much less material but in much greater depth. (vii) Teachers should work to erase the false dichotomy that exists between evolution and religion.

RICHARD F. FIRENZE[1] AND THOMAS O'BRIEN[2]
[1]Biology Department, Broome Community College, Binghamton, NY 13902, USA. [2]Center for Science, Mathematics, and Technology Education, Binghamton University, Binghamton, NY 13905, USA.

Reference

1. R. Firenze, *Rep. Natl. Center Sci. Educ.* **17** (no. 2), (1997).

Keeping an Open Mind?

THE ARTICLE "JUDGE ORDERS STICKERS removed from Georgia textbooks" (C. Holden, News of the Week, 21 Jan., p. 334) quotes a high school science department chairman as being "thrilled" when a sticker asking that evolution be "studied carefully and critically considered" was removed from a textbook by a judge. Pardon me for not being thrilled. The complete wording of the sticker was, "This textbook contains material on evolution. Evolution is a theory, not a fact, regarding the origin of living things. This material should be approached with an open mind, studied carefully and critically considered." Sounds less like "antiscience" and more like good advice to those on all sides of the issue. Oversensitivity to such a sticker indicates a deaf ear and a weak case, not a strong one.

GEORGE ANDERSON
Roseville, MN, USA.

> "Public office does not give the school board… the authority to unilaterally dictate which scientific theory will be believed and which will not."
>
> BROOKS

Keep Censorship Out of Schools

INTELLIGENT DESIGN IS NOT AND CANNOT BE a valid scientific explanation: It does not explain an observable phenomenon in a manner that allows prediction or testing. Evolution does. The scientific community, indeed the community at large, owes Jennifer Miller and her students a large debt for walking out rather than endure political interference in science ("Dover teachers want no part of intelligent-design statement," J. Mervis, News Focus, 28 Jan., p. 505). Public office does not give the school board authoritative powers in science that are not assigned to the most expert scientists, namely, the authority to unilaterally dictate which scientific theory will be believed and which will not. In enunciating their verdict, the school superintendent, Richard Nilsen, demonstrates how easy it is for the political system to abrogate several of our most cherished rights: freedom of speech and freedom of religion.

The school board's action is censorship at best and indoctrination by false statements at worst. Intelligent design is a thinly veiled article of religious faith. Scientists need to put more effort into explaining what science is and how it works.

ALFRED A. BROOKS
Oak Ridge, TN, USA.

Let Students Weigh the Evidence

THE ARTICLE "DOVER TEACHERS WANT NO part of intelligent-design statement" (J. Mervis, News Focus, 28 Jan., p. 505) presents the Dover school board's statement as the work of religious fanatics. The statement (*1*) actually sounds pretty reasonable to me. It says that students should question all theories by evaluating evidence for and against them. This seems to be exactly what science is all about.

Like it or not, intelligent design has adherents. To pretend that the concept of intelligent design does not exist, or to insist that only the "received wisdom" of the scientific community may be presented smacks of the way Galileo and Copernicus were treated when they came up with "outside the box" ideas. Teachers should present the theories and the available evidence, and students should be encouraged to weigh the evidence and then confirm or refute the various theories. This is the essence of the scientific method.

DAVID N. CLARK
Marysville, OH, USA.

Reference

1. See www.dover.k12.pa.us/doversd/cwp/view.asp?A=3&Q=261852.

Don't Dismiss Astrobiology

JEFFREY L. BADA ("A FIELD WITH A LIFE OF ITS own," Books *et al.*, 7 Jan., p. 46) repeats the criticism of astrobiology that the biologist George Gaylord Simpson leveled at exobiology in the pages of *Science* in 1964: "this 'science' has yet to demonstrate that its subject matter exists!" (*1*). In fact, astrobiology is far more than the study of extant extraterrestrial life, but even if that were the sole object of the field, Simpson's criticism must still seem bizarre to many scientists. Much of the most important and compelling research in astronomy, physics, and other fields is exactly concerned with the study of or search for objects or phenomena that may not exist—and that could (and sometimes do) turn out not to exist. Black holes were hypothesized and then searched for long

> "[T]o insist that only the 'received wisdom' of the scientific community may be presented smacks of the way Galileo and Copernicus were treated when they came up with 'outside the box' ideas."
>
> CLARK

Published by AAAS

before compelling evidence regarding their existence accumulated. The same can be said of the search for high-temperature superconductors, proton decay, or the current holy grail of unification, the Higgs boson. Why should we suddenly become giggly when it is biology at stake, rather than physics? In fact, astrobiology merely confronts what is familiar, even commonplace, in many of its sister sciences. There is nothing unique about that, and there is no reason for it to make us uneasy or afraid.

CHRISTOPHER F. CHYBA
SETI Institute and Stanford University, Stanford, CA 94305, USA.

Reference

1. G. G. Simpson, *Science* **143**, 769 (1964).

IN HIS REVIEW OF THE BOOK *THE LIVING Universe* ("A field with a life of its own," Books *et al*., 7 Jan., p. 46), J. L. Bada presents a mistakenly narrow view of astrobiology. Its aim is not just to find life but, rather, to both determine and understand the distribution of life in the universe through time. One extreme possibility is that life exists only on Earth, has never existed anywhere else, and will never be present beyond Earth's orbit. At the alternative extreme, life may have originated on multiple bodies in our solar system and may be ubiquitous beyond. No matter what the answer proves to be, astrobiologists will want to know how the actual distribution of life relates to the occurrence of different planetary environments. Hence, in addition to exploring for evidence of life beyond Earth, astrobiologists study the extreme limits to life, the conditions that make environments habitable, the origin and evolution of life on Earth, the processes responsible for the occurrence of habitable environments in our solar system, and the occurrence of planets and their habitability beyond our solar system.

BRUCE M. JAKOSKY,[1] ARIEL D. ANBAR,[3] DAVID DES MARAIS,[4] DAVID MORRISON,[4] NORMAN R. PACE[2]
[1]Laboratory of Atmospheric and Space Physics, [2]Department of Molecular, Cellular, and Developmental Biology, University of Colorado at Boulder, Boulder, CO 80309, USA. [3]Department of Geological Sciences, Arizona State University, Tempe, AZ 85287, USA. [4]NASA/Ames Research Center, Moffett Field, CA 94035, USA.

Response

THE LETTERS BY JAKOSKY *ET AL.* AND CHYBA demonstrate that my paraphrasing of Simpson's statement about astrobiology (his exobiology) being a field without a known subject elicits the same defensive response it did over 40 years ago (*1*). I certainly agree that the search for life beyond Earth is an important and engaging human endeavor. Simpson's original argument against exobiology as a distinct discipline is that we could learn a lot more about life by studying it right here on Earth, rather than attempting to find life somewhere else. Today, we still do not fully understand how life began on our planet, although there is optimism that this will change in the near future (*2*). It is reasonable to assume that if the conditions that gave rise to life on Earth are widespread, then life must have begun and probably still exists elsewhere. And, of course, we will never know unless we search using the best available tools at our disposal. The fact that we do not yet have any known example of

Letters to the Editor

Letters (~300 words) discuss material published in *Science* in the previous 6 months or issues of general interest. They can be submitted through the Web (www.submit2science.org) or by regular mail (1200 New York Ave., NW, Washington, DC 20005, USA). Letters are not acknowledged upon receipt, nor are authors generally consulted before publication. Whether published in full or in part, letters are subject to editing for clarity and space.

extraterrestrial life makes this challenge all the more exciting but at the same time all the more difficult, especially during this present period of shrinking science budgets.

JEFFREY L. BADA
Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA 92093–0212, USA.

References
1. W. Vishniac, *Science* **144**, 245 (1964).
2. J. L. Bada, *Earth Planet. Sci. Lett.* **226**, 1 (2004).

A Lupus Drug and FDA Approval

WE ARE CONCERNED ABOUT MISINFORMATION in the News Focus article "Lupus drug company asks FDA for second chance" (J. Couzin, 11 Feb., p. 835). It is not true that the U.S. Food and Drug Administration (FDA) has rejected the medication LJP 394, which is under active development by La Jolla Pharmaceutical Company. LJP 394 is currently undergoing testing in an ongoing clinical benefit trial for lupus nephritis that could not occur without the approval and guidance of the FDA. In fact, we understand that this trial is being conducted under a Special Protocol Assessment with the FDA.

It is also not true that the company has requested a "second chance" for the drug. The company has requested consideration of a special development pathway known as accelerated approval. Under this pathway, drugs might be marketed while further development studies are under way. Such an approach has been used numerous times by the FDA for diseases such as cancer and HIV.

JOAN T. MERRILL[1] AND SANDRA RAYMOND[2]
[1]Medical Director, [2]President and CEO, Lupus Foundation of America, Inc., 2000 L Street, N.W., Suite 710, Washington, DC 20036, USA.

Response

AS I REPORTED, LA JOLLA PHARMACEUTICAL Company submitted an application to the FDA detailing over a dozen clinical trials of its lupus drug, LJP 394. Last October, the FDA notified the company that it was declining to approve the drug on the basis of existing data and that an additional, large trial was needed before the drug could be approved. The FDA agreed that a multiyear trial begun the previous summer would suffice. (That trial was originally conceived by the company as a postmarketing trial that would continue while the drug was sold.) The FDA's decision not to approve LJP 394 without further data amounted to a rejection of the company's new drug application.

After the FDA declined to approve LJP 394 without the additional study, La Jolla Pharmaceutical Company asked the FDA for "accelerated approval," which would allow it to market LJP 394 while the trial was taking place. This request, initiated because the company was struggling to pay for costly clinical trials, formed the basis of the article. I believe that my article fairly characterized this appeal as requesting a second chance.

JENNIFER COUZIN

The Norwegian Position on Culling

IN HIS LETTER "FISHERY MANAGEMENT AND culling" (10 Dec. 2004, p. 1890), P. J. Corkeron refers to the White Paper on Norwegian Marine Mammal Policy, which the Norwegian Parliament voted on in May 2003. Corkeron describes the main intent of the policy as a go-ahead for culling marine mammal populations "in the hopes of increased fisheries production."

The central topic of the White Paper is the establishment of an ecosystem-based management regime for marine mammals in areas under Norwegian jurisdiction. This

is a long-term process, and the White Paper proposes steps that can be taken toward this goal. One of these steps is to devise harvesting strategies and propose measures to implement them.

The purpose of the White Paper is to take political action to improve the profitability of whaling, sealing, and the fisheries. Thus, the White Paper talks about profitability, not culling per se. Considerably better profitability, particularly in the sealing industry, is an essential basis for rational and sustainable harvesting of marine mammals within the framework of a future ecosystem-based management regime for living marine resources in Norwegian waters.

Norway's marine mammal policy follows the principle of conservation and sustainable use based on scientific advice. The government also follows the principle that no hunting of seals or whales should be permitted in cases where estimates of stock size are not available.

SVEIN LUDVIGSEN

Norwegian Minister of Fisheries and Coastal Affairs, Ministry of Fisheries and Coastal Affairs, Oslo 0032, Norway.

TECHNICAL COMMENT ABSTRACTS

COMMENT ON "Force-Clamp Spectroscopy Monitors the Folding Trajectory of a Single Protein"

Robert B. Best and Gerhard Hummer

Recent pulling experiments of ubiquitin polyproteins, by Fernandez and Li (Reports, 12 March 2004, 1674), did not show the sharp length contractions expected for cooperative folding of individual protein modules. Simulations of a simple model suggest that the entropic elasticity of the unfolded ubiquitin repeats masks the folding transitions. Intermodule aggregation did not play a significant role in the simulations.

Full text at www.sciencemag.org/cgi/content/full/308/5721/498b

RESPONSE TO COMMENT ON "Force-Clamp Spectroscopy Monitors the Folding Trajectory of a Single Protein"

J. Brujić and J. M. Fernandez

Folding trajectories observed by force-clamp spectroscopy challenge the current view of protein folding with unprecedented results. They reveal that folding is cooperative between the individual domains in the polyprotein chain and takes place only at the end of a trajectory. The simulations by Best and Hummer fail to predict this result and instead explain their continuous folding trajectory by a stochastic folding process.

Full text at www.sciencemag.org/cgi/content/full/308/5721/498c

Clarifications About Teratorns

THE NEWS FOCUS ARTICLE "ANTIEXTINCTION tip: eat to live" (E. Stokstad, 26 Nov. 2004, p. 1466) contains two unfortunate errors. *Teratornis merriami* was cited as "the largest flighted bird ever, with a wingspan of 3 meters or more," but this title actually belongs to the late Miocene teratorn *Argentavis magnificens* of Argentina, which had a wingspan of 6 to 8 m (*1*). Furthermore, contrary to what was inferred about their diet, teratorns were probably not scavengers. They were active predators on small, terrestrial vertebrates (*2*). These errors do not affect the outcome of the study reported, but they perpetuate myths about teratorns.

KENNETH E. CAMPBELL

Department of Vertebrate Zoology, Natural History Museum of Los Angeles County, 900 Exposition Boulevard, Los Angeles, CA 90007–4057, USA.

References

1. K. Campbell, E. Tonni, *Contrib. Sci.* **330**, 59 (1980).
2. K. Campbell, E. Tonni, *J. Vertebr. Paleontol.* **1**, 265 (1982).

Published by AAAS

Comment on "Force-Clamp Spectroscopy Monitors the Folding Trajectory of a Single Protein"

In a recent experiment (*1*), the folding of a single polyprotein consisting of nine N-C linked ubiquitin repeats was monitored under a constant force with an atomic force microscope (AFM). Unlike the discrete, staircase-like increase in the length of the polyubiquitin chain observed in unfolding experiments at higher forces, folding was characterized by a slow contraction in the end-to-end length of the protein (R_{NC}), terminated by a sharp drop to the length at which all the ubiquitin modules are folded. This result has been interpreted as either a breakdown of two-state kinetics under force (*1*, *2*) or evidence of initial aggregation events prior to folding (*3*), because monomeric ubiquitin is known to aggregate at concentrations above 2 μM (*4*). By simulating the dynamics of folding under force using a simplified representation of the protein, we show here that the absence of steplike changes in the overall length upon folding of individual modules can be explained by the elastic properties of the remaining unfolded modules.

In our simulations, the protein is represented by a chain of alpha carbon beads with a Gō-like potential, in which attractive interactions occur only between residues that form contacts in the folded structure (*5*). However, by including favorable interactions for both intra- and intermodule native contacts, folding competes with aggregation. Following the experimental protocol, we started with an unfolded four-module polyubiquitin model under high force and then quenched the force to a low value that favors folding. The simulations capture the experimentally observed behavior (*1*): A rapid initial collapse is followed by a period of slow contraction, characterized by large fluctuations in the protein extension, and a final jump to the fully folded length (Fig. 1A).

What is causing this behavior? In contrast to the experiment in (*1*), we can monitor the folding of each subunit individually in our simulation. The fractions Q_i of native contacts for each subunit i exhibit sharp cooperative transitions from typical unfolded values of Q (<0.5) to a folded Q close to 1 (Fig. 1, B to E). At the same time, the end-to-end distance (between the first and last residue) of each subunit drops to near-native values upon folding, with a concomitant reduction in fluctuations. Remarkably, though, the folding events (red arrows in Fig. 1A) are not at all obvious from monitoring the overall length of the protein. The reason is that the remaining unfolded modules act as soft entropic springs whose large fluctuations mask the decrease in contour length when a single module folds. This effect is not seen in high-force unfolding experiments because the unfolded modules are fully stretched, so the contour length increases in discrete steps upon each unfolding event (*2*). Protein elasticity also explains the slower response for the final jump to the folded state in the AFM experiments (*1*, *3*) compared with recent single-molecule fluorescence studies (*6*).

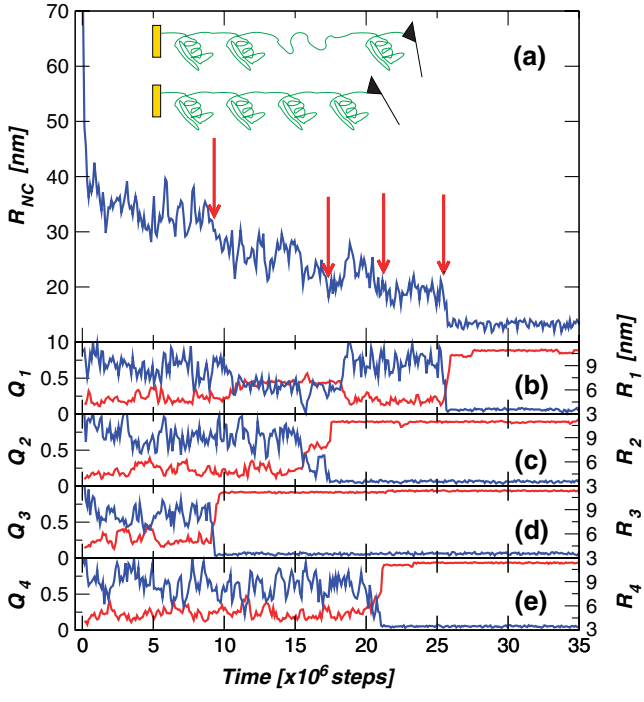


Fig. 1. Refolding simulation of a four-repeat ubiquitin chain under force. The polyprotein is pulled from the termini at a constant low force [inset in (a)], as in the experiment, and (a) the overall protein end-to-end extension (R_{NC}) is monitored over time as the repeats refold. The times at which modules fold are indicated by red arrows in (a); in (b) to (e), the blue traces show the end-to-end length (R_i) and the red traces the corresponding fraction of native contacts (Q_i) for each module i. A value of Q close to 1 indicates that the module is folded.

There is little evidence of aggregation in our simulations, despite the inclusion of native-like interactions between different modules. The fraction of intermodule nativelike contacts between any two modules is always <0.05. Modifying the model to include additional non-native contacts in the energy function resulted in a more frustrated system and slower folding, but did not qualitatively alter the results presented above. In connection with this, we note that ubiquitin is in fact biosynthesized as a polyubiquitin fusion protein similar to the one studied in the AFM experiment (*7*), although the cell is clearly a more complex folding environment.

Although these simulations do not preclude other explanations (which may well contribute in the experiment), they demonstrate that a nontrivial folding model with cooperative folding transitions is sufficient to describe the essential features of the measurements. This interpretation is also consistent with other experimental observations. The increase in folding time with contour length (*1*) can be related to the expected increase in folding time with the number of unfolded repeats for independently folding modules (*8*). In addition, the approximate extrapolation of experimental folding time to zero force, 0.01 s (*1*), is within an order of magnitude of that found using chemical denaturant under similar conditions (*4*).

Individual protein-folding events may be fully resolvable with the experimental setup of Fernandez *et al.* (*1*, *2*) by operating at higher forces whereby fluctuations in the end-to-end length are reduced, although folding is slowed exponentially by force. Alternatively, stiffer linkers could be useful—for example, polyproteins consisting of multiple titin modules, unfolding only at high forces, and a single ubiquitin module.

Robert B. Best *and* ***Gerhard Hummer****
Laboratory of Chemical Physics
National Institute of Diabetes and Digestive and Kidney Diseases
National Institutes of Health,
Building 5, Bethesda MD 20892–0520, USA.
**To whom correspondence should be addressed.*
E-mail: gerhard.hummer@nih.gov

References and Notes

1. J. M. Fernandez, H. Li, *Science* **303**, 1674 (2004).
2. J. M. Fernandez, H. Li, J. Brujic, *Science* **306**, 411c (2004).
3. T. R. Sosnick, *Science* **306**, 411b (2004).
4. H. M. Went, C. G. Benitez-Cardoza, S. E. Jackson, *FEBS Lett.* **567**, 333 (2004).
5. J. Karanicolas, C. L. Brooks III, *Prot. Sci.* **11**, 2351 (2002).
6. E. Rhoades, M. Cohen, B. Schuler, G. Haran, *J. Am. Chem. Soc.* **126**, 14686 (2004).
7. A. M. Weissman, *Nat. Rev. Mol. Cell Biol.*, **2**, 169 (2001).
8. If the folding-time distribution of a single protein module is exponential with mean folding time τ, the mean folding time for an n-module protein is given by $\tau_n = \tau[\gamma + \psi(n + 1)] \approx \tau[\gamma + \ln(n + 1) - (2n + 2)^{-1}]$, where $\gamma \approx 0.5772$ and $\psi(x)$ are Euler's constant and psi function, respectively.

1 November 2004; accepted 24 February 2005
10.1126/science.1106969

Response to Comment on "Force-Clamp Spectroscopy Monitors the Folding Trajectory of a Single Protein"

It is encouraging to see that the protein-folding trajectories observed after a force quench (*1*) have raised interest within the scientific community (*2*). The comment of Best and Hummer accurately points out that the mechanical coupling between the folding domains could mask stepwise folding (*3*). However, their simulations fail to predict our experimental observation that all the domains fold cooperatively at the end of the measured folding trajectories.

In a typical force-quench experiment, a ubiquitin chain is fully unfolded and extended before quenching the pulling force to a low value to trigger folding. The resulting folding trajectories are marked by four distinct phases (*1*, *4*). The first phase is a rapid drop in the length of the unfolded protein associated with elastic recoil (*5*). This phase is followed by a prolonged plateau (phase 2) that implies a search in the conformational energy landscape (*6*), which ends in a faster final contraction in length that can sometimes be resolved into two processes (phases 3 and 4). The folded state of all the domains is reached at the end of phase 4. Fig. 1 shows several of these folding trajectories where the length is scaled by the number of ubiquitin repeats in the chain (from 1 to 7). They are similar in both the shape and amplitude of the changes in length. In particular, the fast contraction in length corresponding to the final transition between the ends of phases 2 and 4 is measured to be 14.6 ± 1.5 nm per protein domain, independent of the stretching force or the number of domains in the chain. It is striking that the folding trajectory of a single ubiquitin (gray trace in Fig. 1) superimposes well with the scaled trajectories of ubiquitin chains containing up to seven repeats. This result can only be explained by a highly cooperative folding between the domains in a chain at the end of the trajectory.

In addition, as we previously reported [figure S4, Supporting Online Material in (*1*)], folding, as determined by the recovery of mechanical stability, was never observed during the long plateau phase 2. Folding was observed in only 7% of the cases during phase 3 and in over 93% after reaching the end of phase 4. This further reinforces the conclusion that the folding of multiple domains is cooperative between the individual domains, with the final chain collapse accompanying the formation of native contacts and folding. By contrast, the trajectories generated by Best and Hummer predict that folding is equally likely anywhere along the pathway, a prediction that fails to reproduce our experimental observations.

Although the shape of the folding trajectories for ubiquitin chains is qualitatively different between our experiments and their simulations, both approaches reveal a complex folding path for individual ubiquitin proteins (*7*). This is highlighted by the well-resolved intermediates that are observed in the single ubiquitin folding trajectories measured with force-clamp spectroscopy (*1*) as well as the single module folding trajectories of Best and Hummer.

The concept of a "two-state" folding reaction has been a useful simplification that allowed generations of biochemists to interpret their observations of folding/unfolding reactions measured from bulk quantities of proteins (*8*). Inevitably, observing folding trajectories at the single-molecule level can no longer be described so simply and therefore necessitates statistical mechanics and a more detailed description of the folding energy landscape (*9*, *10*). In particular, the mechanism by which a protein recovers its folded length after a force quench is still unknown. It is inescapable that the physics of polypeptide collapse under a stretching force will play a crucial role in explaining our trajectories. Unfortunately, a theoretical understanding of these phenomena is lacking. The cooperativity observed at the end of all our folding trajectories is puzzling but clear and cannot be explained by uncorrelated, Markovian models. Explaining this cooperativity will help understand the physical mechanisms underlying protein folding.

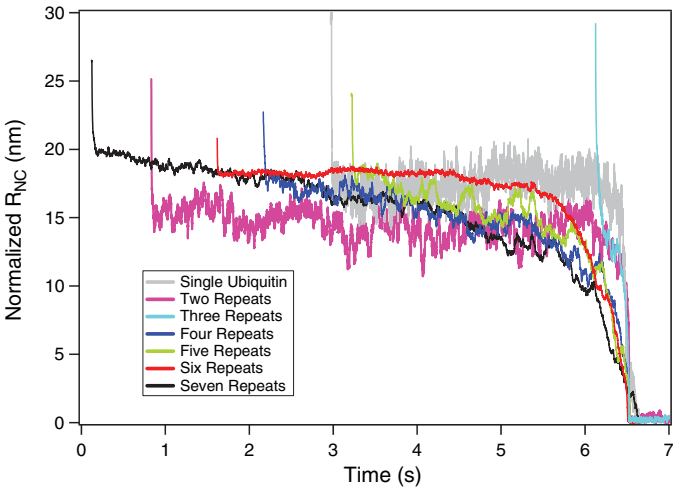


Fig. 1. Normalized folding trajectories of ubiquitin chains measured with force-clamp spectroscopy. The chains vary in length between a single ubiquitin (gray trace) and seven ubiquitin modules (black trace). The folding trajectories have been scaled by the number of ubiquitin repeats in each chain. The normalized overall end-to-end lengths (R_{NC}) show very similar time courses and amplitudes of the final contraction, which underscores the cooperative nature of the observed folding. The origin is set at the time of the force quench for the longest trajectory (black trace); the others are aligned by the final contraction in length for comparison.

J. Brujić *and* ***J. M. Fernandez***
Department of Biological Sciences, Columbia University
New York, NY 10027, USA

References

1. J. M. Fernandez, H. Li, *Science* **303**, 1674 (2004).
2. J. M. Fernandez, H. Li, J. Brujić, *Science* **306**, 411c (2004).
3. R. B. Best, G. Hummer, *Science* **308**, 498 (2005); www.sciencemag.org/cgi/content/full/308/5722/498b.
4. D. Thirumalai, *J. Phys. I* (*Paris*) **5**, 1457 (1995).
5. P. G. De Gennes, *J. Phys. Lett.* **46**, L639 (1985).
6. M. Doi, S. F. Edwards, *The Theory of Polymer Dynamics* (Clarendon, Oxford, 1986).
7. M. Karplus, E. Shakhnovich, in *Protein Folding: Theoretical Studies of Thermodynamics and Dynamics*, T. E. Creighton, Ed. (Freeman, San Francisco, 1992), pp. 127–195.
8. A. R. Fersht, *Structure and Mechanism in Protein Science* (Freeman, New York, 1999).
9. R. L. Baldwin, *Nature* **369**, 183 (1994).
10. T. Lazaridis, M. Karplus, *Science* **278**, 1928 (1997).

29 December 2004; accepted 1 April 2005
10.1126/science.1107675

URBAN POLICY

Learning from the Past to Forge a Future

Peter Muller

Although *Up Against the Sprawl* may be focused on the five-county Los Angeles metropolis that anchors the Southern California region, the development and public policy issues it confronts readily apply to the rest of urban America as well. Metropolitan Los Angeles, of course, has been the hearth of our automobile-suburban culture since at least 1940 (when it gave us the nation's first freeway) and surely merits the "poster child of urban sprawl" label that the editors bestow in their introductory statement. But 21st-century metropolitan Los Angeles is entering a new era as its growth machine, which propelled seemingly endless rounds of urban-landscape expansion over the past six decades, slows amidst a plethora of development-related problems—despite a relentless population influx that over the next 20 years is expected to add yet another 6 million residents to the 16-plus million who already live there.

Up Against the Sprawl
Public Policy and the Making of Southern California
Jennifer Wolch, Manuel Pastor Jr., and Peter Dreier, Eds.
University of Minnesota Press, Minneapolis, 2004. 423 pp. $74.95. ISBN 0-8166-4297-4. Paper, $24.95. ISBN 0-8166-4298-2.

This coming transformation has been thoughtfully anticipated by the book's 20 contributors, all experts on Southern California who spent two years researching the key issues as well as exchanging, integrating, and refining their ideas and chapter drafts. Editors Jennifer Wolch, Manuel Pastor Jr., and Peter Dreier (respectively, a University of Southern California geographer, a University of California, Santa Cruz, economist, and an Occidental College policy scientist) were not only fully involved as authors but also committed to assembling the widest possible range of scholarly perspectives, essential for really doing justice to this multifaceted, interdisciplinary topic. These efforts are organized into a tripartite framework, which is prefaced by a powerful introductory essay that sets the background and defines the emerging issues. Part I traces the region's evolution, emphasizing the infrastructural, industrial, and housing policies that produced the metropolitan geography of race and class. Part II considers the contemporary urban scene and highlights the policy-driven forces that shape population redistribution, housing markets, access to employment opportunities, fiscal dynamics of poverty, and ecological impacts at the metropolitan fringe. Part III turns the focus to the future by examining how certain communities and interest groups are responding to the economic, environmental, and social-equity challenges of urban sprawl.

The basic premise of this collection of studies is that suburban sprawl, intrametropolitan segregation, and urban poverty are connected and that they have been far more heavily shaped by government decisions than by the workings of market forces responding to consumer preferences. Sprawl, therefore, should not be regarded as an accident of public choice: it is mostly the intentional creation of deeply entrenched public policy, a condition that the authors argue can be redirected to produce more efficient and equitable urban spatial patterns. Their case for such change is a strong one, supported by compelling evidence from a number of Southern California sectors and locales, and they sense that a rising tide of public concerns about environmental quality and the negatives of sprawl may be about to provide opportunities for those shifts in policy-making.

A particularly useful discussion is presented by Elizabeth Gearin, whose final-section chapter reviews the smart-growth planning movement. Even though land use–focused smart growth strategies can only rechannel rather than curtail development, they are potentially valuable to metropolitan Los Angeles because they could help steer growth inward and demonstrate the advantages of higher-density living—something the region must regardless soon move toward to accommodate its expected population surge. Gearin also explores two other recently launched planning trends, new regionalism (which highlights the economic interdependence of cities and their suburbs) and urban sustainability (which seeks to establish patterns of development that do not exceed regional carrying capacity). Whereas both might offer more powerful antidotes to sprawl, they also require that public agencies and governments at least begin the politically unpopular redistribution of existing resources. In addition, Gearin contributes to the fascinating final chapter, which surveys seven ongoing, innovative, community-level experiments that could serve as models for regionwide initiatives. These include the introduction of urban-growth limits that can only be extended by direct voter approval; grassroots coalition-building in the form of a bus rider's union, which has favorably impacted metropolitan transit-service provision and policy-making; and an intermunicipal partnership to transform the deindustrializing belt of southeastern Los Angeles County into a corridor of high-tech "gateway cities" linked to the nearby port complex that serves the Pacific Rim.

CREDIT: LANDSAT-7/NASA/USGS

It is often said that wherever urban America is headed, Los Angeles is likely to get there first. By skillfully integrating an array of cutting-edge social and policy science along with perspectives on planning, the editors have given us a splendid overview of the processes, patterns, and issues that mark metropolitan Los Angeles's growth at the dawn of a new era—one that will require the better management of the fabric of 20th-century urbanization that has already stretched this vast region to its geographic limits. Besides the practical lessons the volume offers planners and policy-makers throughout the United States, its analyses of sprawl-related issues constitute a major addition to the academic literature of urban studies and cognate social sciences. Lastly, the book also represents a solid contribution to the public debate about the future of metropol-

The reviewer is at the Department of Geography and Regional Studies, Post Office Box 8067, University of Miami, Coral Gables, FL, 33124–2221. E-mail: pmuller@miami.edu

Published by AAAS

itan growth. The editors are optimistic about that future, which they believe can most effectively be realized by building new political coalitions that combine equity-based regionalism, greater sustainability, and a coordinated smart growth agenda.

10.1126/science.1110467

PSYCHOLOGY

Fleeting Infant Types to Enduring Traits

Paul T. Costa Jr.

Two great questions have captivated the interest of those who would understand how adult personalities come to be. The first is the trajectory of human psychological development. Does it show a pattern of continuous linear growth followed by a period of constancy or greatly reduced change in human characteristics or traits? Or should we award priority to developmental discontinuity, the emergence of qualitatively different characteristics, with developmental change possible at all stages of the human life span? The second enduring question weighs temperaments or biologically based inherited influences (nature) against environmental influences from parenting, peers, and culture (nurture).

Temperaments, introduced by Hippocrates and Galen, were essentially exiled from developmental psychology for most of the 20th century. According to the paramount nurture position, the environment mediates nearly all psychological change and development. A leading exponent of this long-dominant view has been Jerome Kagan, a professor emeritus at Harvard University and one of the world's most celebrated developmental psychologists.

In 1980, Kagan wrote that "There is little firm evidence for the idea that individual differences in psychological qualities during the first two years of life are predictive of similar or theoretically related behaviors a decade hence" (*1*). Rather, he held,

> The child is influenced by experiences from the moment of birth. There is strong evidence for believing that variation in parental practices during the first two years can produce dramatic variation among children in placidity, irritability, hostility, lability, and cognitive capability…The events that fill the years between infancy and adolescence can alter the early dispositions for a great many children. (*1*).

What a change in Kagan's thinking and orientation *The Long Shadow of Temperament* represents. In it, he and Nancy Snidman (the director of Harvard's Infant Laboratory) summarize their extensive longitudinal study of children having one of two extreme temperamental types (identified in early infancy): "inhibited," the high-reactive temperament (shy and timid); and "uninhibited," the low-reactive temperament (behaviorally bold and sociable). The authors hypothesize that the origin of these temperamental biases lies in the differential excitability of a particular limbic structure, the amygdala, extensively studied by neuroscientists and behavioral biologists during the past 20 years. In the book, they report the results of almost two dozen measurements that, they contend, indirectly assess the driving force of amygdala responsiveness.

The Long Shadow of Temperament
by Jerome Kagan and Nancy Snidman
Harvard University Press, Cambridge, MA, 2004. 302 pp. $27.95, £18.95, €25.80. ISBN 0-674-01551-7.

If the claims of Kagan and Snidman pass scientific review, then—finally—one of the great mysteries of human nature will have been solved. How well substantiated are their conclusions? That is not an easy question to answer, especially for the nonspecialist reader. Following the complex and often novel statistical comparisons that the authors use to bolster their case is an arduous task.

The authors cast their findings in almost poetic terms. Their infant high- and low-reactive temperaments "cast long shadows that changed their shapes over the course of development." Similarly, they say that low-reactive infants evolve into adolescents of "sanguine mood" who "experienced delight from new sights, sounds, and conversations that tweaked their understanding of the world." Nonetheless, we must ask how well their hypotheses are supported and whether their integrative attempts have a firm evidentiary foundation.

Kagan and Snidman note that with one exception (the enhanced startle response) the biological data are "in modest accord with the expected outcomes for children who had been high- or low-reactive infants." But there are several troubling inconsistencies in their account. For example, in summarizing their behavioral assessments they comment that "about 1 in 3 high-reactives (22 of 67 children) and 1 in 2 low-reactives (46 of 92 children) had developed social behaviors that were predictable from their infant temperaments. Only 8 high-reactives and 13 low-reactives developed a profile seriously inconsistent with expectations." Seen another way, these data show that two-thirds of the high-reactives do not develop into the inhibited type, and half of the low-reactives do not remain uninhibited. In another example, the authors speculate that the uninhibited profile was better preserved because "family and friends encourage sociability rather than shyness, and American children would rather be sociable than shy." Overgeneralization aside, such environmental sculpting of the sociable disposition is hard to reconcile with their finding that for children evaluated at 4, 7, and 11 years, 61 percent of the high-reactives were always subdued and 33 percent of the low-reactives were always spontaneous.

The evidence presented in *The Long Shadow of Temperament* suggests that infant temperament is a poor predictor of temperament in puberty. In part, these findings may be due to the authors' choice to focus on discrete types, which are statistically and conceptually problematic. Stronger results have come from researchers such as Avshalom Caspi and colleagues (*2*), who conceive of temperament in terms of continuously distributed personality traits. They see moderate continuity from age 3 to age 26, and extensive research shows that personality traits are impressively stable from adulthood through extreme old age (*3*). These traits in turn have been interpreted as adult temperaments (*4*). I hope that Kagan and Snidman, having moved from environmentalists to temperamentalists, continue to advance their thinking and move from discrete types to continuously distributed personality traits.

References

1. J. Kagan, in *Constancy and Change in Human Development*, O. G. Brim, J. Kagan, Eds. (Harvard Univ. Press, Cambridge, MA, 1980), chap. 2.
2. A. Caspi *et al.*, *J. Pers.* **71**, 493 (2003).
3. B. W. Roberts, W. F. DelVecchio, *Psychol. Bull.* **126**, 3 (2000).
4. P. T. Costa, R. R. McCrae, in *Temperament in Context*, T. D. Wachs, G. A. Kohnstamm, Eds. (Erlbaum, Mahwah, NJ, 2001), pp. 1–21.

10.1126/science.1109549

The reviewer is at the National Institute on Aging, National Institutes of Health, Gerontology Research Center, 5600 Nathan Shock Drive, Baltimore, MD 21224, USA. E-mail: ptc@nih.gov

PSYCHOLOGY

The Science of Child Sexual Abuse

Jennifer J. Freyd,[1]* Frank W. Putnam,[2] Thomas D. Lyon,[3] Kathryn A. Becker-Blease,[4] Ross E. Cheit,[5] Nancy B. Siegel,[6] Kathy Pezdek[7]

Child sexual abuse (CSA) involving sexual contact between an adult (usually male) and a child has been reported by 20% of women and 5 to 10% of men worldwide (*1*–*3*). Surveys likely underestimate prevalence because of underreporting and memory failure (*4*–*6*). Although official reports have declined somewhat in the United States over the past decade (*7*), close to 90% of sexual abuse cases are never reported to the authorities (*8*).

CSA is associated with serious mental and physical health problems, substance abuse, victimization, and criminality in adulthood (*9*–*12*). Mental health problems include posttraumatic stress disorder, depression, and suicide (*13*, *14*). CSA may interfere with attachment, emotional regulation, and major stress response systems (*15*). CSA has been used as a weapon of war and genocide and is associated with abduction and human trafficking (*2*).

Much of the research on CSA has been plagued by nonrepresentative sampling, deficient controls, and limited statistical power (*16*). Moreover, CSA is associated with other forms of victimization (*17*), which complicates causal analysis of its role in adult functioning. However, associations in larger scale community and well-patient samples have been confirmed after controlling for family dysfunction and other risk factors (*18*, *19*), in longitudinal investigations that measure pre- and post-CSA functioning (*20*), and in twin studies that control for environmental and genetic factors (*12*, *21*).

Most CSA is committed by family members and individuals close to the child (*1*), which increases the likelihood of delayed disclosure (*22*), unsupportive reactions by caregivers and lack of intervention (*8*, *23*), and possible memory failure [(*24*, *25*), compare (*26*)]. These factors all undermine the credibility of abuse reports, yet there is evidence that when adults recall abuse, memory veracity is not correlated with memory persistence (*27*, *28*). Research on child witness reliability has focused on highly publicized allegations of abuse by preschool operators and has emphasized false allegations rather than false denials (*29*, *30*). Cognitive and neurological mechanisms that may underlie the forgetting of abuse have been identified (*31*–*33*).

Scientific research on CSA is distributed across numerous disciplines, which results in fragmented knowledge that is often infused with unstated value judgments. Consequently, policy-makers have difficulty using available scientific knowledge, and gaps in the knowledge base are not well articulated. We recommend interdisciplinary research initiatives and a series of international consensus panels on scientific and clinical practice issues related to CSA. This can promote (i) increased inclusion of CSA education in the curriculum in medical and mental health fields; (ii) improved education of the public, the media, and professionals who work with alleged CSA victims; (iii) greater visibility and improved dissemination of CSA research; (iv) increased focus on CSA by researchers in a range of disciplines; and (v) improved cost-benefit analyses of intervention, including prevention efforts.

We call on researchers from social science, medical, and criminal justice fields to gather better information on the prevalence (*34*), causes, consequences, prevention, and treatment of CSA. A 1996 report from the Department of Justice (*35*) estimated rape and sexual abuse of children to cost $1.5 billion in medical expenses and $23 billion total annually to U.S. victims. Whereas $2 is spent on research for every $100 in cost for cancer, only $0.05 is spent for every $100 dollars in cost for child maltreatment (*36*). The National Child Traumatic Stress Network is a federally funded network of 54 sites providing community-based treatment to children and their families exposed to a wide range of trauma. The network should be expanded to address the enormous public health consequences of child trauma, and supported to develop new forms of treatment. Even creation of a new Institute of Child Abuse and Interpersonal Violence within the NIH would be justified on the basis of the emotional and economic cost of these problems.

[1]Department of Psychology, University of Oregon, Eugene, OR, 97403–1227; [2]Department of Pediatrics, Children's Hospital Medical Center, Cincinnati, OH 45229; [3] Law School, University of Southern California, Los Angeles, CA 90089; [4]Family Research Laboratory, University of New Hampshire, Durham, NH 03824; [5]Department of Political Science, Brown University, Providence, RI; 02912 [6]NBS Associates, Columbia, MD 21046; [7]Department of Psychology, Claremont Graduate University, Claremont, CA 91711, USA.

*Author for correspondence. E-mail: jjf@dynamic.uoregon.edu

References and Notes

1. D. Finkelhor, *Future Child.* **4**, 31 (1994).
2. World Health Organization (WHO), *World Report on Violence and Health* (WHO, Geneva, 2002); available at www.who.int/violence_injury_prevention/violence/world_report/.
3. R. M. Bolen, M. Scannapieco, *Soc. Serv. Rev.* **73,** 281 (1999).
4. D. M. Fergusson, L. J. Horwood, L. J. Woodward, *Psychol. Med.* **30**, 529 (2000).
5. J. Hardt, *J. Child Psychol. Psychiatry* **45**, 260 (2004).
6. C. S. Widom, S. Morris, *Psychol. Assess.* **9**, 34 (1997).
7. *Child Maltreatment Report 1990* [to 2002] (U.S. Department of Health and Human Services, Washington, DC, 2003); reports from 1995 to 2002 are available at www.acf.hhs.gov/programs/cb/publications/cmreports.htm
8. R. F. Hanson *et al.*, *Child Abuse Neglect* **23**, 559 (1999).
9. C. S. Widom, *Child Abuse Neglect* **18**, 303 (1994).
10. F. W. Putnam, *J. Am. Acad. Child Adolesc. Psychiatry* **42**, 269 (2003).
11. D. M. Fergusson, L. J. Horwood, M. T. Lynskey, *J. Am. Acad. Child Adolesc. Psychiatry* **34**, 1365 (1996).
12. E. C. Nelson *et al.*, *Arch. Gen. Psychiatry* **59**, 139 (2002).
13. B. E. Molnar, S. L. Buka, R. C. Kessler, *Am. J. Public Health* **91**, 753 (2001).
14. B. E. Molnar, L. F. Berkman, S. L. Buka, *Psychol. Med.* **31**, 965 (2001).
15. M. D. De Bellis *et al.*, *J. Clin. Endocrinol. Metab.* **78**, 249 (1994).
16. J. Briere, *J. Consult. Clin. Psychol.* **60**, 196 (1992).
17. J. G. Noll *et al.*, *Interpers. Violence* **18**, 1452. (2003).
18. C. L. Battle *et al.*, *Personal. Disord.* **18**, 193 (2004).
19. R. Roberts, T. O'Connor, J. Dunn, J. Golding, ALSPAC Study Team, *Child Abuse Neglect* **28**, 525 (2004).
20. S. Boney-McCoy, D. Finkelhor, *J. Consult. Clin. Psychol.* **64**, 1406 (1996).
21. S. Dinwiddie *et al.*, *Psychol. Med.* **30**, 41 (2000).
22. D. W. Smith *et al.*, *Child Abuse Neglect* **24**, 273 (2000).
23. D. M. Elliott, J. Briere, *Behav. Sci. Law* **21**, 261 (1994).
24. J. J. Freyd, *Betrayal Trauma* (Harvard, Cambridge, MA, 1996).
25. J. J. Freyd, A. P. DePrince, E. L. Zurbriggen, *J. Trauma Dissoc.* **2** (3), 5 (2001).
26. G. Goodman *et al. Psychol. Sci.* **14**, 113 (2003)
27. C. J. Dalenberg, *J. Psychiatry Law* **24**, 229 (1996).
28. L. M. Williams, *J. Trauma. Stress* **8**, 649 (1995).
29. S J. Ceci, M. Bruck, *Jeopardy in the Courtroom: A Scientific Analysis of Children's Testimony* (American Psychological Association, Washington, DC, 1995).
30. T. D. Lyon, *Cornell Law Rev.* **84**, 1004 (1999).
31. M. C. Anderson *et al.*, *Science* **303**, 232 (2004).
32. A. P. DePrince, J. J. Freyd, *Psychol. Sci.* **15**, 488 (2004).
33. H. Sivers, J. Schooler, J. J. Freyd, in *Encyclopedia of the Human Brain*, V. S. Ramachandran, Ed. (Academic Press, San Diego, CA, 2002), vol. 4, pp. 169–184.
34. For example, the Bureau of Justice Statistics collects data on crimes against people aged 12 and older.
35. T. R. Miller, M. A. Cohen, B. Wiersema, *Victim Costs and Consequences: A New Look* (U.S. Department of Justice, Washington, DC, 1996).
36. F. W. Putnam, in *The Cost of Child Maltreatment: Who Pays?* K. Franey, R. Geffner, R. Falconer, Eds. (Family Violence and Sexual Assault Institute, San Diego, CA, 2001), pp. 185–198.

10.1126/science.1108066

APPLIED PHYSICS

How to Build a Superlens

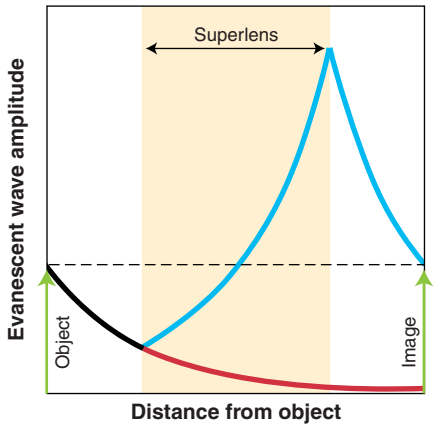
David R. Smith

In 1968, Veselago conceived of a material whose index of refraction, unlike that of any known material, could be negative (*1*). He suggested that this negative-index material would reverse nearly all known optical phenomena. Amid considerable initial skepticism, negative refraction was experimentally confirmed in an artificially structured material at microwave frequencies in 2001 (*2*). The work prompted a flurry of activity by researchers to further explore and demonstrate the properties of negative-index materials.

One of the most dramatic—and controversial (*3*)—predictions to emerge from this activity was a speculation by Pendry (*4*) that a thin negative-index film should function as a "superlens," providing image detail with a resolution beyond the diffraction limit to which all positive-index lenses are subject. On page 534 of this issue, Fang *et al.* (*5*) confirm the theoretical predictions of Veselago and Pendry. They show that a planar negative-index lens can indeed produce a sharp image by virtue of a new mechanism: evanescent wave refocusing.

Conventional positive–refractive index lenses require curved surfaces to bend the rays emanating from an object to form an image. Yet, Veselago noted that negative–refractive index lenses are not subject to the same constraint. He found that a planar slab of material with a refractive index of –1 could also produce an image (*1*). For this lens, diverging rays from a nearby object are negatively refracted at the first surface of the slab, reversing their trajectories so as to converge at a focus within the material. The rays diverge from this focus and are again negatively refracted at the second surface, finally converging to form a second image just outside the slab. Although it produces an image, the planar lens differs from conventional curved-surface lenses in that it does not have an optical axis, does not focus parallel rays, and has a magnification that is always unity.

On careful reexamination of this planar lens, Pendry found that the ray picture applied by Veselago did not tell the whole story (*4*). The electromagnetic field of an object includes not only propagating waves, but also near-field "evanescent" waves that decay exponentially as a function of distance away from the object. The evanescent waves carry the finest details of the object, but cannot be recovered by conventional positive-index lenses, which can therefore resolve objects to no better than roughly one-half of the illuminating wavelength—the diffraction limit.



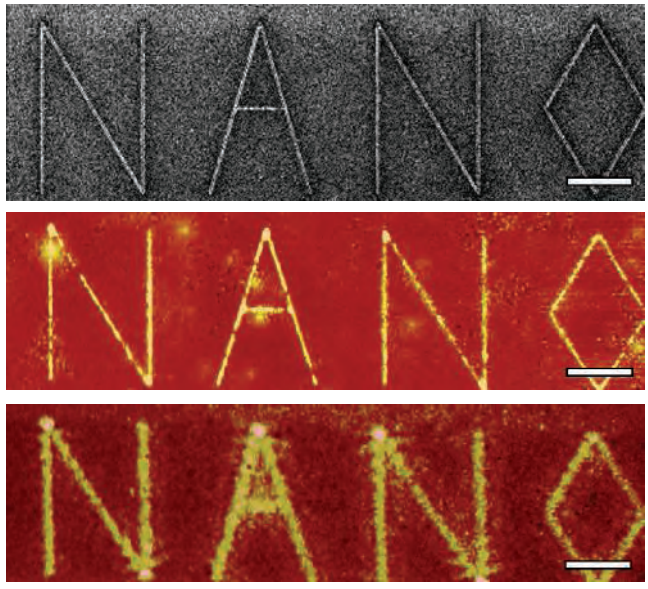
The principle of evanescent wave refocusing. The exponentially decaying wave from the object on the left grows exponentially within the planar negative-index lens (blue curve). On the other side of the lens, it decays again until it has reached its original value at the image plane. These components of the object are lost in the absence of the negative-index lens (red curve).

Pendry found that a planar negative-index slab should refocus the evanescent waves, at least to some extent. An evanescent wave decaying away from an object grows exponentially across the planar negative-index lens (see the first figure). On exiting the lens, the wave decays again until it reaches the image plane, where it has the same amplitude with which it started. Unlike any other lens, the resolution limit of the planar negative-index lens is determined by how many evanescent waves from the object can be recovered, rather than by the diffraction limit.

There is no theoretical limit on the resolution of the superlens, but for a reasonable amount of evanescent wave refocusing to occur, the distances between the object, its image, and the slab surfaces—and the thickness of the slab itself—must all be small relative to the wavelength. If these conditions are not met, the evanescent waves from the object decay to the extent that their recovery becomes impractical owing to material losses and other material imperfections of the lens.

This constraint, as it turns out, also hides a virtue. A negative-index material requires both the electric permittivity ε and the magnetic permeability μ to be less than zero. At optical wavelengths, there are no known materials that have a negative μ; this would appear to rule out a superlens at optical wavelengths. However, over scales much less than a wavelength, electric and magnetic effects decouple, and only one of the two parameters has to be negative. Because $\varepsilon < 0$ occurs naturally in silver and other metals at visible wavelengths, a thin metallic film can act as an optical superlens.

Following Pendry's suggestion, Fang *et al.* now demonstrate evanescent wave refocusing in the context of optical lithography. In the experiments, a thin film of silver



A demonstration of evanescent wave refocusing. Fang *et al.* show that evanescent wave refocusing can be used to create the optical image (**center**) of a lithographically written object (**top**) with subwavelength resolution. Without the lens, the image resolution is much lower (**bottom**). Scale bar, 2 μm.

CREDIT: SECOND FIGURE FROM (*5*)

The author is in the Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708, USA. E-mail: drsmith@ee.duke.edu

Published by AAAS

serves as the superlens that transfers the image of a lithographically written pattern to a nearby layer of photoresist. But coaxing evanescent waves to grow requires two stringent criteria to be satisfied. First, the surface of the film must be extremely smooth; otherwise, surface imperfections scatter the incident light and wash out the finer details carried by the evanescent waves. Second, the thickness of the silver film must be optimized: If it is too thick, material losses dominate over the evanescent wave refocusing, and none of the information carried by the evanescent waves is recovered in the image. The film produced by Fang *et al.* meets both criteria, with an optimal thickness of ~35 nm and a surface roughness of less than 1 nm (*6*).

The demonstration of superlensing requires a subwavelength object. In the experiments of Fang *et al.*, such an object is formed by the light that passes through thin slits (with a width of 40 nm) that have been patterned into an otherwise opaque chromium mask. Because the slits are narrow relative to the wavelength (365 nm), the light is strongly diffracted, with most subwavelength features being contained in the evanescent waves. As a result, the image blurs rapidly as a function of distance away from the mask. The reduction in image quality is noticeable over a distance of tens of nanometers, as can be seen in the second figure.

Fang *et al.* use the light that passes through the chromium mask and the lens to expose a layer of photoresist, where the optical image is converted into a topographic map of peaks and valleys that can be scanned with an atomic force microscope. As an example, the authors patterned the word "NANO" into the mask (see the second figure, top panel). In the absence of the silver superlens, the lines that form the letters are diffuse (bottom panel), with a measured line width of more than 300 nm. With the silver superlens, the evanescent waves are recovered, and markedly better resolution is obtained (middle panel), with an observed line width of less than 90 nm.

The results of Fang *et al.* (*5*) confirm that the predicted phenomenon of evanescent wave refocusing is indeed possible at visible wavelengths. This important advance not only resolves a controversial aspect of negative-index materials, but also opens the door to a variety of possible applications, including higher resolution optical imaging and nanolithography. Optical elements can now be designed to access and exploit the near-field of light.

References

1. V. G. Veselago, *Sov. Phys. Usp.* **10**, 509 (1968).
2. R. A. Shelby, D. R. Smith, S. Schultz, *Science* **292**, 77 (2001).
3. D. R. Smith, J. B. Pendry, M. C. K. Wiltshire, *Science* **305**, 788 (2004).
4. J. B. Pendry, *Phys. Rev. Lett.*, **85**, 3966 (2000).
5. N. Fang, H. Lee, C. Sun, X. Zhang, *Science* **308**, 534 (2005).
6. Z. Liu, N. Fang, T.-J. Yen, X. Zhang, *Appl. Phys. Lett.*, **83**, 5184 (2003).

10.1126/science. 1110900

NEUROSCIENCE

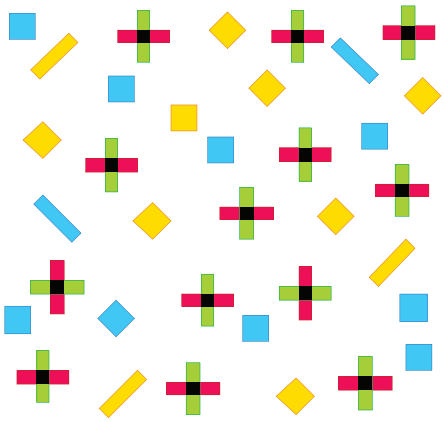
Watching Single Cells Pay Attention

Jeremy M. Wolfe

As visual organisms, we spend much of our time engaged in visual search behavior. We seek to make the current object of our desire into the current object of our visual attention and motor action. You want a sip of coffee. There is the mug. Then you wonder, where is the "%" sign on the keyboard? Next, the ring of the phone redirects your attention to that object. Most searches such as these go by so quickly and effortlessly that we don't notice the search aspect at all. We do notice when the task becomes more difficult: Where is that corkscrew in the kitchen gadget drawer? Ah, there it is, in full view, but somehow not noticed until after a prolonged period of searching. Insights into how area V4 of the visual cortex might participate in these sophisticated search tasks are revealed by Bichot *et al.* (*1*) on page 529 of this issue.

So, how do we carry out these search tasks? Behavioral and physiological experiments conducted over more than a quarter century have emphasized one of two types of mechanism: parallel processing, in which all (or many) objects are analyzed at once (*2*, *3*); and serial processing, in which one (or very few) of the available objects are selected for specialized analysis (*4*, *5*).

You may be able to get a qualitative appreciation for these modes of processing by searching for one of the objects in the figure. Find the blue diamond. You will probably notice that all of the blue items seem to make themselves available to you at the same time. If you now search for the yellow square, the blue items recede into the background, while the yellow ones take center stage. Obviously, the stimulus has not changed. Your search goal has changed your analysis of that stimulus. If you are asked to search for the plus sign with red-vertical and green-horizontal elements, all the red and green plus signs may seem to become salient. But at the same time, you may be aware that some scrutiny of single items is needed before you find the plus sign having red linked to vertical. (If it felt instantaneous, go find the *other* plus sign with a red-vertical element. There are two.) The color and orientation features seem to be present almost immediately, but the binding of a color to an orientation seems to require something more.



Finding a needle in a haystack. Your analysis and experience of this display will change depending on whether you are looking for a blue diamond or for a plus sign with a red-vertical element. Bichot *et al.* reveal how different aspects of attention modulate the response of neurons in area V4 of the visual cortex as monkeys perform similar tasks (*1*).

Here, then, are two rather different types of processing that might be seen to fall into the general category of "attention." First, it seems possible to attend to a distributed set of items based on features like color. And second, it seems possible to select individual items for fixation or to select an item for further analysis even if it is not fixated. In most, if not all, search tasks, these processes interact to produce an effective visual search (*6*, *7*). Parallel information about features will guide your serial selection of individual objects—as you pick your favorite bits out of a fruit salad, for example.

The author is at the Visual Attention Laboratory, Brigham and Women's Hospital and Harvard Medical School, Cambridge, MA 02139, USA. E-mail: wolfe@search.bwh.harvard.edu

In their new study, Bichot *et al*. (*1*) demonstrate that both sorts of processing occur in area V4 of the visual cortex during a single search task. They recorded neuronal responses from area V4 in the brain of alert macaque monkeys who were searching through displays of colored shapes similar to those shown in the figure. In the first experiment, at the start of a trial, the monkey would be shown a stimulus that cued him to the target color (for example, red) or shape (for example, star). To find this target, the monkey typically made a succession of eye movements that changed the position of the visual stimuli on the retina. During a trial, Bichot *et al*. recorded from a V4 neuron that was sensitive to stimulation in one specific region of the retina (its "receptive field"). Beyond being sensitive to one location in space, a V4 neuron might also have a preferred color and/or shape. Thus, as the monkey moved his eyes around, he presented different stimuli to the receptive field of the neuron under study. If the neuron preferred red stimuli, then, by definition, that neuron responded more vigorously whenever the monkey saw red. Of more interest, on trials when the target color was red, the neuron produced a still larger response. This happened even when the monkey was not about to make a quick eye movement to the target. The red item had not yet become the specific object of attention, but the response of the neuron still received a boost because red was the desired color. Moreover, neurons that preferred the target feature synchronized their activity, perhaps giving them a better chance of activating subsequent postsynaptic neurons.

So much for the parallel enhancement of all items on the basis of a feature like color. What about the selection of specific items? Imagine the situation in which the monkey is searching for a red item, and a red item lies in the receptive field of the studied neuron. We know from previous work (*8*) that covert attention shifts to an object before it is fixated by the eyes. So Bichot *et al*. went back over their data and sorted the responses into two categories: responses from just before the monkey made an eye movement toward the red item and responses from just before he made an eye movement somewhere else. They found that the neuron responded more strongly just before the eyes fixated on the red item. Thus, it seems that the act of attentional selection that precedes serial fixation also enhanced the response of the neuron. What we have here is attractive evidence at the level of single cells indicating that parallel feature processes are guiding serial selection of plausible targets for further scrutiny.

Understanding how monkeys (and presumably other primates, such as ourselves) perform search tasks is of more than academic interest. Visual search is a task each of us performs a thousand times a day, from searching for a coffee cup to looking for a face in a crowd. However, some searches are more important than others. As a society, we have created many artificial but critically important search tasks, such as airport baggage screening and routine mammography. Many of these tasks are complicated and currently performed imperfectly. We eagerly await development of new ways to improve human performance on such tasks or the invention of machines that could take over or assist with them. Understanding how biological systems do so well at performing a range of search tasks should help us to improve the outcome of those artificial search tasks on which we, quite literally, stake our lives.

References

1. N. P. Bichot, A. F. Rossi, R. Desimone, *Science* **308**, 529 (2005).
2. J. Palmer, *Curr. Dir. Psychol. Sci.* **4**, 118 (1995).
3. J. Duncan, G. W. Humphreys, *Psychol. Rev.* **96**, 433 (1989).
4. E. Bricolo, T. Gianesini, A. Fanini, C. Bundesen, L. Chelazzi, *J. Cogn. Neurosci.* **14**, 980 (2002).
5. G. F. Woodman, S. J. Luck, *J. Exp. Psychol. Human Percept. Perform.* **29**, 121 (2003).
6. J. M. Wolfe, *Psychon. Bull. Rev.* **1**, 202 (1994).
7. J. M. Wolfe, *Trends Cogn. Sci.* **7**, 70 (2003).
8. E. Kowler, E. Anderson, B. Dosher, E. Blaser, *Vision Res.* **35**, 1897 (1995).

10.1126/science.1112616

CELL BIOLOGY

A Fishing Buddy for Hypothesis Generators

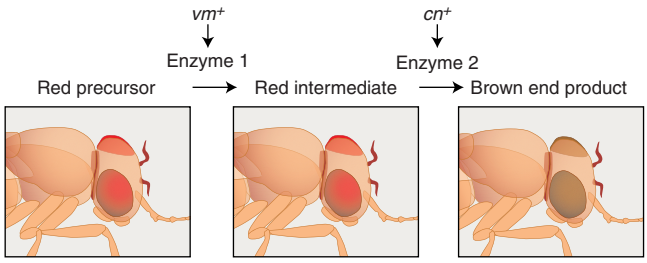
Roger Brent and Larry Lok

In the last half of the 20th century, the dominant experimental modality in biology was hypothesis-directed research. Of course, biology has a proud tradition of important insights arising from undirected poking around and following hunches. However, until genomic biology made undirected fishing for information more respectable (*1*), the most common response to requests for money for such projects was dismissal with the term "fishing expedition." The study by Sachs *et al*. (*2*) on page 523 of this issue suggests it may be time to reexamine this prejudice.

In their study, Sachs, Nolan, Lauffenburger, and their co-workers outline what may be a powerful new way to fish. They combine measurements of different signal transduction proteins in large numbers of individual human $CD4^+$ T lymphocytes and computational frameworks called Bayesian networks with experimental perturbations that are close to hypothesis-free. These investigators not only regenerated known causal relationships among the signaling proteins but also predicted new connections that they verified by targeted testing. For example, they predicted interpathway cross-talk between the Erk1 and Akt kinases. Their approach suggests a way that "fishing expeditions" and investigator reasoning might supplement one another, generating testable assertions about chains of causation and action in biological systems.

The authors are at the Molecular Sciences Institute, Berkeley, CA 94704, USA. E-mail: brent@molsci.org, lok@molsci.org

Consider how biologists identify candidate gene products and suggest how these might act in chains of cause and effect. With "genomic" methods, two grounds for identification correspond to logical fallacies: "guilt by association" (for example, these two proteins touch one another, and might cooperate in the same cellular process) and "post hoc ergo propter hoc" (for example, this gene regulator is expressed before this transcript appears and therefore might regulate production of that transcript) (*3*). In contemporary biology, such observations are supplemented by additional information including DNA sequences, which can suggest direction of action from the biochemical function of the encoded proteins and similarities to known pathways in other organisms. Such inferences are complemented by experimental methods that more rigorously establish flows of action and consequence. These approaches go back to the work of Ephrussi and Beadle in the 1930s. These experimenters isolated fruit flies with mutations in genes for eye color. They then showed that an eye disk containing the *cinnabar* (*cn*) gene product but lacking the *vermillion* (*vm*) gene product produced a wild-type eye when transplanted into flies that lacked *cinnabar* but contained *vermillion*; they also showed that the converse was not true (*4*, *5*) (see the top figure). This phenomenon, called "epistasis," established both that *cn* and *vm* act in an eye-color "pathway" and that the wild-type *vm* gene product must act first in order for the wild-type *cn* gene product to exert its effect.



Order of action in eye color. Cause and effect relationships established by epistasis (*4*, *5*). The vm^+ gene product from the recipient acts on red precursor pigment to produce a red intermediate, and the cn^+ gene product from the eye disk acts on the intermediate to produce wild-type brown pigment. Flies with mutations in the *vermillion* (vm^-) and *cinnabar* (cn^-) genes have red eyes. An eye disk from a $vm^- cn^+$ larva transplanted into a $vm^+ cn^-$ host results in wild-type brown eyes, but the converse transplantation results in red eyes, due to the enzyme pathway depicted. An eye disk from a $vm^- cn^+$ donor receives the vm^+ product from a $vm^+ cn^-$ recipient. The vm^+ product from the recipient acts on red precursor pigment to produce substance A, and the cn^+ product from the disk acts on A to produce B, the wild-type pigment.

Sachs *et al*. (*2*) provide real progress in developing descendant methods to assign gene products to pathways and to learn the order of their actions. This progress rests on their ability to measure the amounts of various proteins (in this case signaling proteins) in large numbers of individual cells. Nolan and co-workers have spent years perfecting techniques for targeting proteins in (fixed) lymphocytes with fluorescent antibodies and for measuring the amounts of proteins by flow cytometry (*6*). Their work culminated last year in a beautiful paper (*7*) that quantified different protein species involved in signal transmission in lymphocytes from patients with acute myeloid leukemia. These investigators stimulated the leukemic cells by various means and then classified the leukemias according to the differences in their complements of signal transduction proteins. They proposed that this method of classification could eventually guide mechanism-based treatment.

A second element necessary for this progress is the development of Bayesian statistical methods to formalize common concepts of causation, work begun by Judea Pearl (*8*, *9*). Bayesian networks compute and display relationships of statistical dependency between system variables. Variables are represented by nodes in a graph. Arrows between nodes represent some dependencies between the corresponding variables explicitly; other dependencies are left implicit. If arrows point from "cause" nodes to "effect" nodes, the network is called a causal Bayesian network. The form of such graphical models is familiar to geneticists, but the difference is that Bayesian network methods automatically generate numerous alternative models from experimental data and assign probabilities to each different network structure and each different arrow. The methods initiated by Pearl take into account measurements after experimental perturbations of the process to obtain causal Bayesian networks. In these networks, the adage "correlation does not imply causation" is still true, but insights from perturbations provide additional information that allows, *inter alia*, determination of which correlations arise from causal relationships, and the direction each arrow of causality points. The third element is the ability to carry out interventions that initiate the process and perturb its function. These three capabilities enabled Sachs *et al*. to perturb human $CD4^+$ T lymphocytes in different ways in order to collect data from thousands of cells subjected to each trial, and to infer casual relationships among the measured variables.

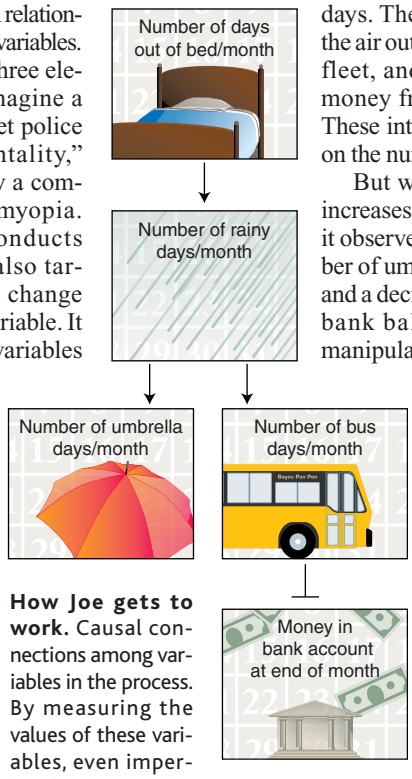
To illustrate how the three elements work together, imagine a city dominated by a secret police agency, "the Instrumentality," which is distinguished by a combination of power and myopia. The Instrumentality conducts grand operations, and also targeted interventions that change the value of almost any variable. It can perturb many more variables than it can measure, and its measurements are often imprecise.

The Instrumentality has become interested in the process by which an inhabitant, Joe, goes to work (see the bottom figure). Joe is clinically depressed, but, after sunrise each morning, if he is not too depressed, he gets out of bed. When the weather is good he walks to work. When it is raining, he uses his umbrella as he walks to the bus stop, and then rides the bus. By the end of the month, riding the bus depletes his bank balance. The Instrumentality can measure values for all of these variable aspects of Joe's life, and, because it can make the sun come up in the morning, it can perform as many trials as it wants. Note that the values of some variables are conventionally correlated even though they are not causally connected. For example, the number of days Joe carries an umbrella to work each month correlates positively with the number of days he rides the bus, and negatively with his bank account at the end of the month, although taking the umbrella doesn't cause him to ride the bus. Even without perturbing the system, the Instrumentality could posit from these observations that rain days caused umbrella days and bus days, or that rain days, umbrella days, and bus days all had a common cause.

The Instrumentality learns more by intervening in downstream steps of the process. For example, it can decrease the number of umbrella days to zero by hiding Joe's umbrella, or can increase the number to 30 by gluing the umbrella to his hand. Whatever the number of umbrella days, Joe will continue to take the bus on rainy days and walk on sunny ones; so the number of bus days will be unaffected by the number of umbrella days. The Instrumentality can let the air out of the tires of the city bus fleet, and deposit and withdraw money from Joe's bank account. These interventions have no effect on the number of umbrella days.

But when the Instrumentality increases the number of rain days, it observes an increase in the number of umbrella days and bus days, and a decrease in his end-of-month bank balance. By contrast, no manipulation of umbrella days or bus days affects the number of rain days. Thus, by taking measurements after interventions, the Instrumentality reveals relationships between the observables of Joe's life that correspond to the chain of causation in our natural-language narrative, and, for some variables, which way the arrows point: in this case, from rain days to umbrella days and not the other way around.

Four more points about this example merit mention. First, in many cases, causal independence can be established by observation alone. For example, although the Instrumentality vigilantly tracks variables related to the purchase



How Joe gets to work. Causal connections among variables in the process. By measuring the values of these variables, even imperfectly, for most months of the year, an outside observer can establish correlations among them. By using Bayesian network methods together with "non–hypothesis-directed" perturbations to the values of these variables, the observer can identify causal connections and assign directionality to them. Given enough trials, the methods can operate efficiently on "noisy" data. In most biological systems, application of these methods is limited by the kinds of variables researchers can quantify and the number of times the researchers can repeat each set of measurements.

CREDITS: PRESTON HUEY/*SCIENCE*

of water skis, these show independence from Joe-related variables under all conditions, and drop out of consideration. Second, the Bayesian network approach allows the observer to start with a hypothesis about the structure of a network, and to perform targeted (as opposed to blundering) experiments to test particular network structures. Third, the particular approach used by Sachs *et al*. fails to identify cases in which downstream events feed back into upstream events (for example, when Joe's bank balance hits zero, he cannot take the bus). However, coupling measurements of variables at different time intervals with "dynamic Bayesian networks" may allow identification of feedback relationships. Finally, existing methods cannot identify causal connections between variables the Instrumentality does not know exist. In this example, the probability that Joe gets out of bed is influenced by whether he has filled a prescription for an antidepressant drug at the nearby drugstore the month before. Thus, Joe's antidepressant purchases seem to be relevant upstream "causal" input for the number of bus days. But if the Instrumentality has not yet learned about antidepressants and drugstores, it will not be able to discover the additional causal link.

When we return from the Instrumentality to our own world, we find that biologists are very good at making targeted perturbations. In genetically tractable organisms, performing these perturbations often depends on making the right mutant. In cell lines and in less tractable organisms, perturbations might be better effected by RNA interference, "protein genetic," or (for people) pharmacological approaches. We also see that for Bayesian network methods to realize their promise, researchers will need to get much better at measuring relevant variables. For intracellular events, variables include but are not limited to, numbers of regulatory molecules, modified molecules, and specific molecular complexes, and the percent occupation of regulatory sites upstream of genes. To be useful, measurement methods will need to operate on individual cells, or, at the very least, to allow large enough numbers of trials to yield causal assertions reliable enough to merit further experimental testing.

The Sachs *et al*. work is important because it suggests how researchers might develop a package of capabilities to enable systematic fishing expeditions. Such a package would generate testable ideas based on (relatively) error-prone high-throughput measurements made after (relatively) uninformed experimental treatments and could help experimenters refine those ideas after quick tests. Such capabilities seem well suited to one of the grand challenges of 21st- century biology: the grouping, ordering into pathways, and description of function for the numerous weakly acting and incompletely penetrant genes that quantitatively modify important phenotypes in humans and other organisms.

References and Notes

1. R. Brent, *Cell* **100**, 169 (2000).
2. K. Sachs, O. Perez, D. Pe'er, D. A. Lauffenburger, G. P. Nolan, *Science* **308**, 523 (2005).
3. R. Brent, *Curr. Biol.* **9**, 338 (1999).
4. G. W. Beadle, B. Ephrussi, *Genetics* **21**, 225 (1936).
5. B. Ephrussi, G. W. Beadle, *Genetics* **22**, 479 (1937).
6. P. O. Krutzik, G. P. Nolan, *Cytometry* **55A**, 61 (2003).
7. J. M. Irish *et al.*, *Cell* **118**, 217 (2004).
8. J. Pearl, in *Probabilistic Reasoning in Intelligent Systems* (Morgan-Kauffman, San Mateo, CA, 1988).
9. J. Pearl, *Causality: Models, Reasoning, and Inference* (Cambridge Univ. Press, Cambridge, UK, 2000).
10. We thank the Alpha Project for support.

10.1126/science.1110535

PLANT SCIENCES

Recognition at a Distance

Paul Schulze-Lefert and Stéphane Bieri

A key step in the evolution of eukaryotic immune systems was the ability to discriminate between self and nonself. Evidence suggests that animals and plants independently evolved dedicated and highly variable receptor families for recognition of nonself structures. The outcome of interactions between plants and the pathogenic microbes that invade them largely relies on a repertoire of receptors that serve as a radar system for detecting pathogen-derived nonself molecules. The function and specificity of these receptors were originally defined by genetic studies. Such studies revealed that for plants to recognize their intruders and to mount an effective resistance response, there needed to be a match between a strain-specific pathogen effector and its corresponding plant host resistance (*R*) gene product (*1*). Detection of a pathogen effector by a plant R receptor frequently leads to rapid death of plant host cells at sites of attempted invasion as part of the immune response. Most known *R* genes encode intracellular receptors containing a nucleotide binding site and leucine-rich repeats (LRRs) or membrane-bound surface receptors containing extracellular LRRs (*2*). Two new studies—by Coaker *et al*. (*3*) on page 548 of this issue and by Rooney *et al*. (*4*) in this week's *Science Express*—describe encounters between pathogen-secreted effector molecules and their host targets in *Arabidopsis* and the tomato (*Lycopersicon*), respectively. Although this interorganismal molecular liaison has entirely different consequences for the effector target proteins, in both cases, their manipulation holds the key to a better understanding of how plant immune receptors recognize nonself.

Many *Arabidopsis* ecotypes contain the plasma membrane–associated intracellular R protein RPS2 (see the figure). This protein specifically detects and mounts an immune response to strains of the bacterial pathogen *Pseudomonas syringae*, which produce the AvrRpt2 effector protein. AvrRpt2 is delivered into the plant cytosol by a specialized bacterial secretion system and is cleaved near its amino terminus. The carboxyl-terminal cleavage product is sufficient to trigger the RPS2-dependent immune response and is predicted to adopt a secondary structure typical of a cysteine protease (*5*). Although attempts to detect direct interactions between RPS2 and AvrRpt2 have been unsuccessful, both proteins physically associate with the *Arabidopsis* protein RIN4. A complex between RPS2 and RIN4 is constitutively present in healthy (unchallenged) plants, but RIN4 disappears when AvrRpt2 is delivered into plant cells. Importantly, mutations in any of three amino acid residues in the carboxyl terminus of AvrRpt2 (predicted to be essential for catalytic activity of the putative *Pseudomonas* protease) disrupts the processing of AvrRpt2, the RPS2-dependent immune response, as well as elimination of RIN4 (*5*–*7*). This finding prompted the proposal that RPS2 might recognize the result of AvrRpt2's proteolytic activity, that is, the removal of RIN4.

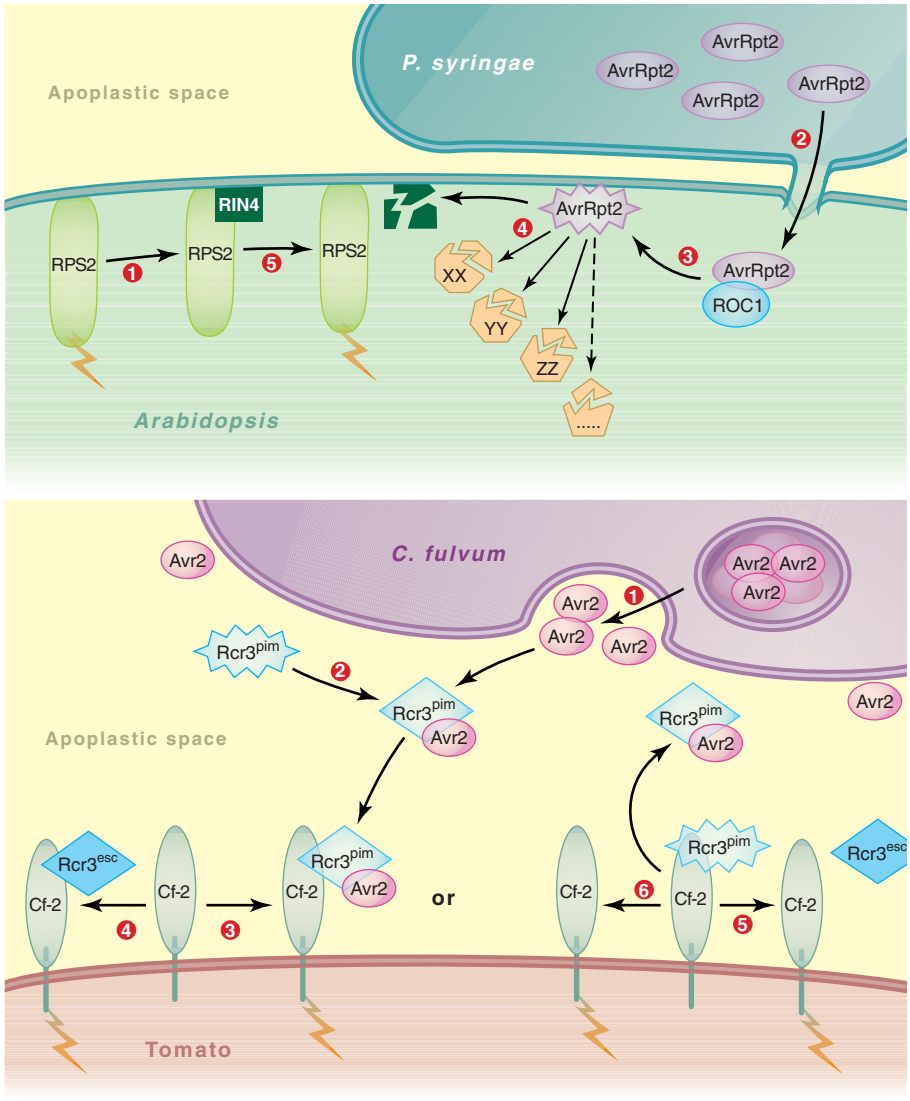
Coaker *et al*. started from the puzzling observation that processing of AvrRpt2 could be detected in all eukaryotic but not prokaryotic extracts tested, including those from *P. syringae*. This observation implies the existence of a eukaryotic cofactor required for AvrRpt2 processing. Using a combined biochemical and genetic approach, the authors identified this cofactor as a single-domain cyclophilin, a folding catalyst that facilitates cis/trans isomerization of prolyl bonds. Cyclophilin activity is required for proper AvrRpt2 self-cleavage, and this in turn may be a critical step for the correct subcellular localization of

P. Schulze-Lefert is in the Department of Plant Microbe Interactions, Max-Planck-Institut für Züchtungsforschung, D-50829 Köln, Germany. E-mail: schlef@mpiz-koeln.mpg.de S. Bieri is at the Institute for Plant Biology, University of Zürich, 8008 Zürich, Switzerland.

AvrRpt2 in plant host cells. Identifying the cyclophilin-dependent self-cleavage activity of AvrRpt2 provided the first indication that RIN4 might be a direct substrate of the bacterial protease. A similar stretch of amino acids surrounding AvrRpt2's self-processing site is also present in two regions within RIN4. Indeed, recombinant full-length AvrRpt2 and the single-domain cyclophilin were necessary and sufficient to specifically cleave RIN4 at these two sites in vitro.

These findings are biologically relevant because substitutions of RIN4 residues at its carboxyl-terminal cleavage site abolish both RIN4 elimination and RPS2 activation in plants (*8*). The authors of this study also determined whether RIN4 cleavage results in a loss of physical association with RPS2, or merely derepresses RPS2 activity while maintaining physical contacts with RIN4 cleavage products. They used in vivo coexpression of DNA constructs encoding RIN4 cleavage products and RPS2. Activity of RPS2 in the presence of RIN4 cleavage products indicated that in vivo release of RPS2 from its RIN4 partner is essential for triggering the immune response. This probably explains previous findings showing autoactivation of RPS2 in *rin4* mutant plants in the absence of the pathogen. Collectively, these data strongly favor a model in which RIN4 negatively regulates RPS2 activity. This mode of regulation permits indirect activation of the AvrRpt2-dependent R receptor through proteolytic elimination of RIN4.

This indirect intracellular perception of a *Pseudomonas* effector protein may be analogous to extracellular recognition of the Avr2 effector protein of the fungus *Cladosporium fulvum* by the *Cf-2 R* gene product of tomato (see the figure), the subject of a complementary paper by Rooney *et al*. (*4*). Cf-2 is a transmembrane receptor–like protein with extracellular LRRs. Previous genetic analysis of Cf-2-mediated resistance revealed that a papain-like protease of the tomato plant, Rcr3, in the extracellular leaf space is required for Cf-2 activity (*9*). Rooney *et al*. showed that this protease is not a signaling component of the Cf-2–triggered immune response but rather is crucial for Cf-2–dependent recognition of the Avr2 fungal effector protein. To monitor Rcr3 protease activity during coimmunoprecipitation experiments, the authors used a biotinylated "suicide" substrate that irreversibly and covalently binds and inhibits the active site of the protease. They demonstrated that Avr2 specifically associates with and inhibits Rcr3 protease activity in the tomato plant in vivo and also after heterologous expression in yeast. Heterologously synthesized Rcr3 or Avr2 or the suicide substrate–locked form of Rcr3 all failed to trigger a Cf-2–dependent immune response when injected into *rcr3* mutant tomato plants containing wild-type *Cf-2*. In contrast, Cf-2 was specifically activated when Rcr3-Avr2 complexes were injected into tomato leaves containing mutant *rcr3* and wild-type *Cf-2*. This finding and the existence of an autoactive *Rcr3* allele (*9*), which activates Cf-2 in the absence of Avr2, suggests that a conformational change in Rcr3 imposed by binding of Avr2 or mimicked by the autoactive Rcr3 allele is the trigger for Cf-2 activation. Whether Cf-2 activity is negatively regulated by binding to Rcr3 in healthy plants—analogous to the negative regulation of RPS2 by RIN4 in



A subtle approach to resisting invasion. Indirect recognition of pathogen effector proteins by plant immune receptors in *Arabidopsis* (**top**) and the tomato (**bottom**). (Top) (1) RIN4 binds to and suppresses the activity of the R receptor RPS2 in healthy plants. (2) The bacterial effector protein AvrRpt2 is secreted by *P. syringae* into host plant cells by a specialized secretion system. (3) *Arabidopsis* cyclophilins (for example, ROC1) interact with AvrRpt2 and activate this bacterial effector protein. The AvrRpt2 protease becomes localized to the plasma membrane after autoproteolytic cleavage of its secretion signal peptide. (4) Activated AvrRpt2 then cleaves RIN4 and at least three other *Arabidopsis* proteins (*16*). (5) RIN4 cleavage products can no longer suppress RPS2 activity. (Bottom) (1) The fungal effector protein Avr2 of *C. fulvum* is secreted into the apoplastic space surrounding host tomato plant cells. Avr2 associates with the secreted plant protein $Rcr3^{pim}$ (an allele originating from wild tomato *Lycopersicon pimpinellifolium*), and inhibits its protease activity. (2) There is a binding-dependent conformational change in Rcr3. (3) Altered Rcr3 then binds to the extracellular R receptor Cf-2, which is activated by indirect recognition of Avr2. (4) An autoactive $Rcr3^{esc}$ variant (originating from *L. esculentum*) may be a conformational mimic of the $Rcr3^{pim}$ state enforced by Avr2 binding. (5) Alternatively, $Rcr3^{pim}$ (but not $Rcr3^{esc}$) may inhibit Cf-2–triggered immune responses. (6) In this case, conformational changes in $Rcr3^{pim}$ enforced by Avr2 binding sequester $Rcr3^{pim}$ away from Cf-2, thereby derepressing the R receptor.

CREDIT: KATHARINE SUTLIFF/*SCIENCE*

Published by AAAS

Arabidopsis—is not known because validated *rcr3*-null mutant plants are not available (*9*). Thus, we can infer two possible modes of Cf-2 activation (see the figure). Secretion of Avr2 during pathogenesis may sequester Rcr3 away from constitutive Cf-2–Rcr3 complexes, thereby derepressing Cf-2 activity. Alternatively, formation of Avr2–Rcr3 complexes may trigger a conformational change in Rcr3, enabling it to bind to and activate Cf-2. In either case, Cf-2–dependent recognition of Avr2 is likely to be indirect, taking place without physical interaction between the fungal effector protein and the plant host R protein.

Work on other plant resistance responses mediated by pairs of host resistance and pathogen effector proteins supports an indirect mode of nonself recognition (*10, 11*). Of particular note is the recognition of the *P. syringae* effector AvrRpm1 by the intracellular and plasma membrane–associated RPM1 receptor of *Arabidopsis*. Both proteins were found to physically associate with *Arabidopsis* RIN4 rather than interacting directly with each other. Thus, RIN4 appears to be a host target for multiple *Pseudomonas* effector proteins (*11*). However, RIN4 does not disappear upon delivery of AvrRpm1 into plant cells. The exact biochemical alteration in RIN4 mediated by AvrRpm1 is poorly understood, but a change in RIN4 phosphorylation seems likely to be involved in RPM1 activation (*11*). An indirect mode of recognition appears to be the common theme in these cases, and clearly plant immune receptors are capable of recognizing biochemically diverse alterations of effector targets, including phosphorylation status, proteolytic cleavage, and conformational changes.

Indirect recognition of nonself in plants is an elegant alternative solution to direct nonself recognition by the adaptive immune systems of vertebrates. Vertebrates evolved dedicated somatic recombination systems for the generation of receptor diversity and specialized immune cells to recognize any potentially harmful nonself molecular structures (*12*). Indirect pathogen detection in plants appears to be as effective as direct nonself recognition in vertebrates. However, fewer receptors are needed—for example, there are only ~120 nucleotide binding LRR-type receptors in the *Arabidopsis* genome (*13*)—and specialized immune cells are not required. R protein–mediated surveillance of only those host protein assemblies that are critical for successful invasion by parasites may have been an important step in helping plants, with their limited set of receptors, to survive. Indeed, it is conceivable that *Arabidopsis* RIN4 and tomato Rcr3 are virulence targets. However, the roles of these two host proteins in cellular reprogramming during pathogenesis remain mysterious. In addition, although conformational changes in R proteins are likely to be critical for their activation (*14*), we do not have detailed insights into this process owing to a lack of R protein crystal structures. Such structures might help to identify the immediate targets of activated R proteins, which are as yet unknown. Finally, it will be important for future studies to compare the current findings with the recognition mechanics of a second nonself perception system in plants, the so-called PAMP (pathogen-associated molecular pattern) receptors. These receptors detect conserved pathogen-derived molecular structures present in multiple microbial species, such as a peptide derived from the bacterial motor protein flagellin (*15*) .

References

1. H. Flor, *Annu. Rev. Phytopathol.* **9**, 275 (1971).
2. J. L. Dangl, J. D. G. Jones, *Nature* **411**, 826 (2001).
3. G. Coaker, A. Falick, B. Staskawicz, *Science* **308**, 548 (2005); published online 3 March 2005 (10.1126/science.1108633).
4. H. C. E. Rooney *et al.*, *Science*, 21 April 2005 (10.1126/science.1111404).
5. M. J. Axtell *et al.*, *Mol. Microbiol.* **49**, 1537 (2003).
6. M. J. Axtell, B. J. Staskawicz, *Cell* **112**, 369 (2003).
7. D. Mackey *et al.*, *Cell* **112**, 379 (2003).
8. B. Day *et al.*, *Plant Cell*, **17**,1292 (2005).
9. J. Krüger *et al.*, *Science* **296**, 744 (2002).
10. F. Shao *et al.*, *Science* **301**, 1230 (2003).
11. D. Mackey *et al.*, *Cell* **108**, 743 (2002).
12. M. Kasahara *et al.*, *Trends Immunol.* **25**, 105 (2004).
13. B. C. Meyers *et al.*, *Plant Cell* **15**, 809 (2003).
14. P. Moffett *et al.*, *EMBO J.* **21**, 4511 (2002).
15. C. Zipfel *et al.*, *Nature* **428**, 764 (2004).
16. S. T. Chisholm *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 2087 (2005).

10.1126/science.1111725

APPLIED PHYSICS

Toward a Universal Memory

Johan Åkerman

When it comes to computers, mp3 players, digital cameras, and other electronic gadgets, there is no such thing as too much memory. Whether it is more Flash memory for taking high-resolution digital pictures, a bigger hard drive for digital video files, or more random access memory (RAM) to view them on the computer, the appetite for ever more memory at ever-increasing densities appears insatiable. An emerging technology, magnetoresistive RAM, promises additional functionality and improved memory performance that will enable yet more applications and open up system designs that are not possible today.

Today's dominant solid-state memory technologies—static RAM, dynamic RAM, and Flash—have been around for a long time, with Flash the youngest at 21 years (*1*). Their longevity can be explained in part by mutually beneficial differentiation. Each technology does a single thing very well, but many systems need all three memory types to deliver overall good performance at reasonable cost. However, the gain from differentiation comes at the cost of increased system and fabrication complexity, particularly in so-called embedded applications, where an entire electronic system is implemented on a single chip with static RAM, dynamic RAM, and Flash often used side-by-side.

All three technologies have advantages and disadvantages. Static RAM has excellent read and write speeds, integrates readily into the process technology of embedded applications, and requires little power for data retention. However, its large cell size (a typical memory bit requires six transistors) makes it impractical for embedded applications that require a lot of memory. Embedded static RAM is used for cache memory in microprocessors, where high speed is more important than large amounts of memory.

Dynamic RAM uses a single transistor and a storage capacitor per cell and thus provides a denser architecture than static RAM, at the expense of increased embedded-process complexity. Because the stored charge tends to leak out of the capacitor, dynamic RAM requires constant power to refresh its bit state every few milliseconds. Because of its high power consumption, large amounts of dynamic RAM are impractical for portable electronics with limited battery life.

In contrast to static and dynamic RAM, Flash memory offers nonvolatile data storage; that is, its information is not lost when the power is turned off. Nonvolatility is highly desirable in portable electronics, because nonvolatile data retention does not consume any battery power. Flash also has high density and moderately fast read access time, but its write mode is too slow and its write endurance far too limited for many applications. In addition, embedded Flash needs its own high-voltage drivers, complicating the design and manufacturing process.

The author is with Freescale Semiconductor, 1300 North Alma School Road, Chandler, AZ 85224, USA. E-mail: johan.akerman@freescale.com

For some time, researchers have tried to devise nonvolatile alternatives to Flash. The goal is a "universal memory" that combines the best attributes of static RAM, dynamic RAM, and Flash. Such a memory would eliminate the need for multiple memories in many applications, would improve system performance and reliability by avoiding data transfer between multiple memories, and would reduce overall system cost.

Magnetoresistive RAM (MRAM) is currently the most promising contender for a memory with such universal characteristics. It combines nonvolatility with relatively high read and write speeds and unlimited endurance. Furthermore, the MRAM storage element resides in the metal interconnect layers, well above the silicon, allowing its process to be optimized independently from the underlying semiconductor process. MRAM is therefore cost effective to integrate and is ideally suited for embedded-memory applications.

State-of-the-art MRAM combines a magnetoresistive magnetic tunnel junction with a single-pass transistor for bit selection during read (see the figure) (*2*). The tunnel junction has a free magnetic layer, a tunneling barrier, and a fixed magnetic layer. The magnetization of the fixed layer is prevented from rotating, whereas the magnetization orientation of the free layer can be switched and is used for information storage. The resistance of the tunnel junction depends on the relative magnetization orientation of the free layer with respect to the fixed layer. For tunnel junctions with a NiFe free layer and an aluminum oxide tunneling barrier, the maximum difference in resistance (the magnetoresistance) is about 40 to 50%.

To read a single bit, the transistor is turned on, a bias of ~0.3 V is applied to the bit, and the memory state is determined by measuring the amount of current that flows through the bit. Programming is achieved by passing current through two perpendicular write lines, one above and the other below the selected bit; these are respectively termed the "bit line" and the "digit line" (see the figure). The lines are clad with magnetic material to focus the applied field to the bit for improved write selectivity and increase the field magnitude by a factor of ~2, reducing the write power consumption by a factor of ~4.

MRAM faces several challenges before it can be introduced to the market on a large scale. The first challenge relates to the switching current distribution. The two-dimensional (2D) write selection scheme outlined above requires tight and uniform switching current distributions. Successful programming of the selected bit requires that the combined write field from both write lines must be greater than the bit's switching field. In addition to the selected bit, thousands of bits along the two write lines (called 1/2-selected bits) see 71% of the write field. As a consequence, a 4-Mb memory will suffer 1/2-select disturbs, unless the standard deviation of the bit-to-bit switching current due to inevitable material and processing variations is less than 6%. Because the 1/2-select disturb process is thermally activated, the actual distribution must be even tighter to ensure proper operation over the life of the memory (typically 10 years).

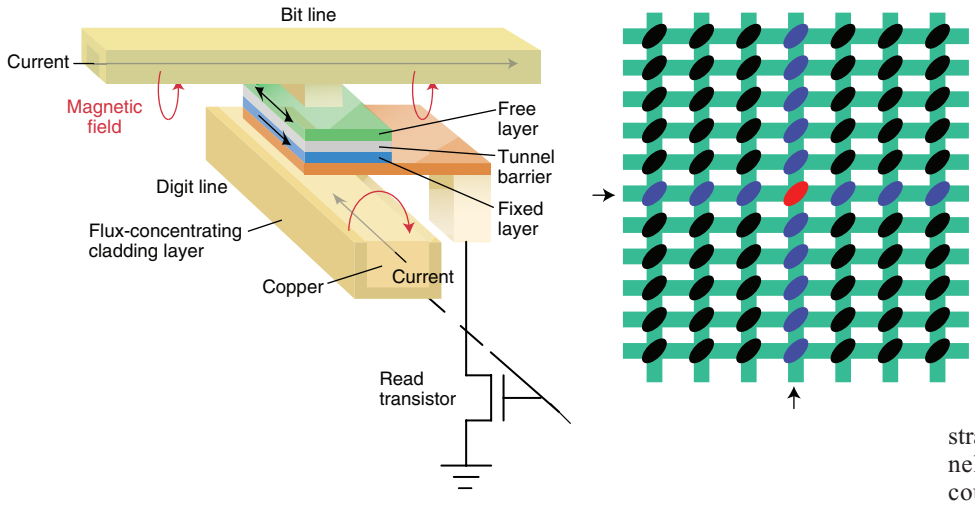
As a solution to the 1/2-select problem, the late Leonid Savtchenko of Motorola proposed a novel free-layer structure and a phase-sensitive write pulse scheme, in which the free layer rotates rather than switches (*3*). A full rotation takes place only if a bit sees both field pulses; as a result, 1/2-selected bits are less susceptible to thermal activation than are unselected bits. Because the free-layer rotation toggles the bit state, unipolar write currents can be used, further simplifying the design. A necessary pre-read before write effectively reduces the number of write pulses by 50%. This toggle-write method is used in the 4-Mb MRAM under development at Freescale Semiconductor (*4*).

A different approach is taken by Cypress Semiconductor, whose 256-kb lower density MRAM avoids the 1/2-select problem altogether by having individual write transistors for each bit. This design can also reduce the overall write current, but it comes at the cost of increased cell size.

The second concern for MRAM is its relatively small readout signal, which effectively limits its read speed. The available signal is roughly proportional to the magnetoresistance divided by the bit-to-bit resistance distribution. IBM has obtained read times of 3 ns (3×10^{-9} s) in 1-kb research memories, Freescale has demonstrated a 25-ns cycle time for its 4-Mb MRAM (*5*), and Cypress Semiconductor is targeting a 70-ns cycle time for its 256-kb MRAM (*6*). However, a magnetoresistance of more than 230% was recently demonstrated in junctions that use MgO as the tunneling barrier (*7*, *8*). Use of this material could lead to much faster MRAM operation, provided the resistance distributions are as tight as for aluminum oxide barriers.

As with any new technology, customers will worry about its long-term reliability. The most obvious concerns relate to the long-term stability of the ultrathin tunneling barrier, the stability of the magnetic layers in the free layer, and data retention. Accelerated tests show that these mechanisms have negligible impact on memory performance at operating conditions.

The tunneling barrier is likely to be highly stable, because aluminum oxide has a high breakdown voltage even at very small thicknesses, MRAM uses a low operating voltage, and only the magnetic tunnel junctions that are being read are subject to any voltage stress. Accelerated dielectric breakdown studies indeed show barrier lifetimes far greater than 10 years (*9*). Interdiffusion between the magnetic layers may affect the switching performance over time, but accelerated tests indicate that over 10 years of use, virtually no change in switching performance at operating temperatures would occur. MRAM data retention is inversely proportional to the thermal flip rate of the free layer, but at present bit dimensions, accelerated tests predict no observable



2D write selection with MRAM. (Left) MRAM bit cell with a magnetic tunnel junction in series with a transistor for bit read selection. Perpendicular write lines above and below the magnetic tunnel junction select a single tunnel junction during programming. **(Right)** Top view of an MRAM array, highlighting the fully selected bit (red) in the center and 1/2-selected bits (blue) along each current-carrying write line. In toggle-MRAM, all bits are oriented at 45° with respect to the write lines.

Published by AAAS

thermal flip rate at operating conditions.

Future generations of MRAM will use smaller tunnel junctions and will thus have to readdress the above challenges. Going toward smaller dimensions must not introduce more bit-to-bit variations or jeopardize data retention. The switching current will not substantially increase with reduced bit size (provided that other dimensions, such as the proximity to the write lines and their width, also decrease). But the current density will scale inversely with the conductor area, and electromigration may therefore become an issue. At that point, spin momentum transfer (*10*)—switching by a spin-polarized current through the bit—might become a viable alternative to 2D write selection.

This year, Cypress Semiconductor became the second company (after Freescale Semiconductor) to announce that it has shipped fully functional MRAM samples to potential customers. Many other companies have demonstrated multi-Mb MRAM prototypes. It is now only a matter of time before the first volume shipments of MRAM devices take place.

References

1. F. Masuoka, M. Asano, H. Iwahashi, T. Komuro, *IEEE IEDM Tech. Digest*, 464 (1984).
2. M. Johnson, Ed., *Magnetoelectronics* (Elsevier, Oxford, 2004).
3. L. Savtchenko, B. N. Engel, N. D. Rizzo, M. F. Deherrera, J. A. Janesky, U.S. Patent 6,545,906 B1, 8 April 2003.
4. B. N. Engel *et al.*, *IEEE Trans. Magn.* **41**, 132 (2005).
5. For the Freescale Semiconductor data sheet for MR2A16A, see www.freescale.com and enter Part Number MR2A16A.
6. For the Cypress Semiconductor datasheet for CY9C62256, see www.chipcatalog.com/Cypress/CY9C62256-70SC.htm.
7. S. S. P. Parkin *et al.*, *Nat. Mater.* **3**, 862 (2004).
8. S. Yuasa *et al.*, *Nat. Mater.* **3**, 868 (2004).
9. J. Åkerman *et al.*, *IEEE Trans. Dev. Mat. Rel.* **4**, 428 (2004).
10. J. A. Katine *et al.*, *Phys. Rev. Lett.* **84**, 3149 (2000).

10.1126/science.1110549

CELL BIOLOGY

Guiding ATM to Broken DNA

Robert T. Abraham and Randal S. Tibbetts

DNA damage poses a continuous threat to genomic integrity in mammalian cells. To cope with this problem, these cells have evolved an elaborate network of sensor, transducer, and effector proteins that coordinate cell-cycle progression with the repair of the initiating DNA lesion. A particularly lethal form of DNA damage is the DNA double-strand break (DSB). The cellular response to DSBs must be swift and decisive—requirements that are capably fulfilled by a serine-threonine kinase in the nucleus called ATM (ataxia-telangiectasia mutated). This nuclear protein serves as a key signal transducer in the DSB response pathway. ATM is a member of the phosphoinositide 3-kinase related kinase (PIKK) family, which includes several important proteins required for genome surveillance (*1*). Humans that lack functional ATM suffer from a devastating syndrome called ataxia telangiectasia (AT), characterized by cerebellar neurodegeneration, premature aging, immunodeficiency, extreme sensitivity to radiation, and heightened susceptibility to developing cancer. The severe pathologies associated with AT are attributable largely, if not entirely, to defective DSB recognition and repair. Exposure to ionizing radiation or other DSB-inducing agents triggers a prompt increase in ATM kinase activity, suggesting that ATM is a proximal transducer of DNA damage signals (*1*). On page 551 of this issue, Lee and Paull (*2*) offer new insights into the molecular mechanism that relays damage signals from DNA to ATM.

Enhanced online at www.sciencemag.org/cgi/content/full/308/5721/510

R. T. Abraham is in the Signal Transduction Program, The Burnham Institute, La Jolla, CA 92037, USA. E-mail: abraham@burnham.org R. S. Tibbetts is in the Department of Pharmacology, University of Wisconsin-Madison, Madison, WI 53706, USA. E-mail: rstibbetts@wisc.edu

Seminal studies 2 years ago by Bakkenist and Kastan (*3*) revealed that, in undamaged cells, ATM resides as a catalytically inactive dimer or higher order multimer. DNA damage induced by ionizing radiation triggers the auto- or trans-phosphorylation of the serine amino acid residue at position 1981 (Ser^{1981}) in the ATM polypeptide. This leads, in turn, to the dissociation of inactive ATM complexes into catalytically active ATM monomers. The authors made the striking observation that nearly the entire nuclear pool of ATM molecules was phosphorylated on Ser^{1981} within minutes of cellular exposure to low doses of ionizing radiation that induced only a few DSBs. To explain this highly efficient amplification mechanism, the authors proposed that even a single DSB causes a far broader alteration in chromatin structure that encompasses megabase regions of genomic DNA. This creates a suitable platform for the prompt activation of hundreds of ATM dimers after DSB induction. Consistent with this epigenetic model for ATM activation, the authors demonstrated that treatment of cells with chromatin-disrupting agents provoked widespread phosphorylation of ATM under conditions that did not produce detectable DSBs.

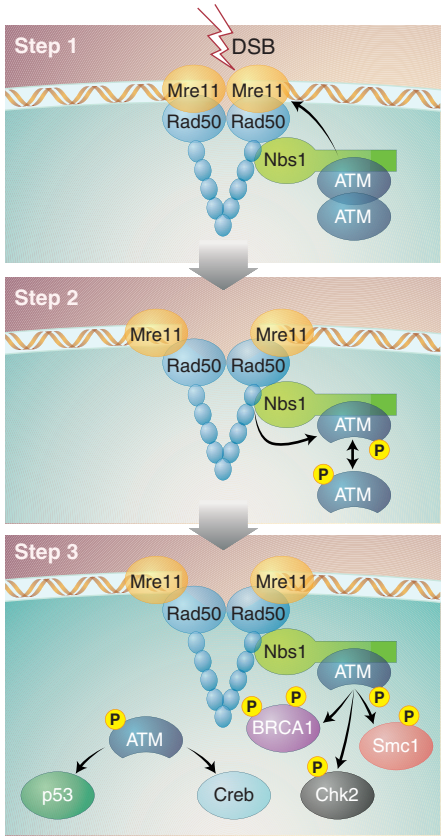
The new study by Lee and Paull (*2*) highlights the Mre11-Rad50-Nbs1 (MRN) complex as an essential mediator of ATM recruitment to and activation by DSBs. The MRN complex has a long history of association with the ATM-dependent checkpoint pathway (*4*). Hypomorphic mutations in the *NBS1* and *MRE11* genes give rise to Nijmegen breakage syndrome (NBS) and an AT-like disorder (ATLD), respectively. The clinical features of ATLD are indistinguishable from those of AT, whereas NBS patients (and cells from these patients) display a somewhat attenuated version of the AT phenotype. Mre11 is a DNA binding protein that possesses 3′,5′-exonuclease activity, as well as an endonuclease activity that cleaves DNA hairpins. Rad50 is a member of the structural maintenance of chromosomes (SMC) family. It forms homodimers that associate with two Mre11 molecules to yield tetrameric Mre11-Rad50 (MR) complexes (see the figure). The two arms of the MR complex allow this structure to form bridges between free DNA ends or between sister chromatids. The contribution of the Nbs1 subunit to the MRN complex is not well understood, although numerous studies show that Nbs1 expression is required for optimal phosphorylation of ATM substrates in cells damaged by ionizing radiation. Bakkenist and Kastan (*5*) recently argued that the partial ATM signaling defects observed in cells from NBS patients indicate that Nbs1 positively influences, but is not essential for, activation of ATM. However, as these authors point out, many NBS cells express a truncated form of Nbs1 that contains an intact carboxyl terminus. It turns out that this region of Nbs1 binds directly to ATM and is important for recruitment of ATM to sites of DNA damage (*6*). Thus, the hypomorphic *NBS1* allele expressed in NBS cells may mask an obligate role for Nbs1 in ATM activation. Earlier findings indicated that the Nbs1 partner protein, Mre11, is equally indispensable for ATM activation after DNA damage (*7*, *8*). Lee and Paull (*2*) now offer compelling biochemical evidence to support the conclusion that the MRN heterotrimer both recruits and activates ATM at DNA damage sites.

Lee and Paull (*9*) earlier demonstrated that the protein kinase activity of purified

human ATM was stimulated by recombinant MR or MRN complexes in the absence of DNA. In these biochemical assays, ATM activation was monitored by phosphorylation of an exogenously added substrate, such as p53. The authors speculated that the binding of MR or MRN to ATM enhanced access of substrate to the ATM kinase domain. However, the lack of a requirement for DNA in these assays hinted that the purification procedure may have favored the isolation of "preactivated" ATM monomers. In their latest study, the authors carefully isolated dimeric (or oligomeric) and monomeric populations of ATM molecules (*2*). With ATM dimers as the starting point, both DNA and the complete MRN complex were required for stimulation of ATM kinase activity. The MR subcomplex was responsible for recognition of the free ends of DNA duplexes, and for Rad50-mediated unwinding of the DNA ends to generate single-stranded DNA (ssDNA), an essential step leading to the recruitment and dissociation of ATM dimers. Interestingly, the ATM-related kinase, ATR, also recognizes ssDNA as a marker for DNA lesions that interfere with progression of the replication fork (*10*). Thus, the accumulation of ssDNA appears to provide a common signal for the recruitment of ATM and ATR to different types of DNA damage. The conservation of recruitment strategies for these two protein kinases apparently extends beyond the configuration of the DNA target. Jackson *et al*. (*6*) recently showed that an evolutionarily conserved motif at the carboxyl terminus of Nbs1 and ATR-interacting protein (ATRIP) mediates direct interactions with the respective partner kinases of these proteins, ATM and ATR. This interaction is required for efficient recruitment of the PIKKs to sites of DNA damage.

The most surprising outcome of the experiments by Lee and Paull (*2*) is that mutation of Ser^{1981} to an alanine (Ala) residue that cannot be phosphorylated has no effect on either the ATM dimer-to-monomer transition, or the stimulation of ATM kinase activity by MRN and DNA. This result is sure to engender considerable debate in the field, given that the Bakkenist and Kastan model (*3*) highlights Ser^{1981} phosphorylation as the pivotal event leading to ATM activation. The discrepant findings reported by Lee and Paull (*2*) may simply reflect the fact that their reconstitution assay offers only a partial view of the mechanism of ATM activation in intact cells. In the biochemical assay, the concentrations of purified DNA and MRN may be sufficiently high to drive the ATM dimer-monomer equilibrium toward the monomeric state, even in the absence of Ser^{1981} phosphorylation. In the intact cell, however, ATM phosphorylation may be critical to prevent the rapid reassociation of ATM monomers at DSB sites. A more trivial explanation is that the recombinant ATM (Ser^{1981}→Ala) "dimers" prepared by Lee and Paull are oligomeric complexes containing cell-derived ATM molecules bearing the Ser^{1981} phosphorylation site. Phosphorylation of one such site in an oligomeric complex may be sufficient to induce complex disassembly and monomer formation in the biochemical assay.

The new studies by Lee and Paull (*2*) focus attention on the MRN complex as both a sensor and effector of ATM activation and signaling in response to DSBs (see the figure). Complementary results from Jackson and co-workers (*6*) indicate that the Nbs1 subunit serves as a bridge between ATM and the DNA-bound MR heterodimer. Where do these new findings leave us with respect to the Bakkenist and Kastan model for ATM activation (*3*)? The results of Lee and Paull are consistent with the idea that the conversion of ATM dimers to monomers is a key event during ATM activation. However, the importance of Ser^{1981} phosphorylation for ATM activation is now open for debate and further experimentation. The generation of mice bearing Ser^{1981}→Ala mutations in both *ATM* alleles should provide definitive evidence for or against the original proposal (*3*) that Ser^{1981} phosphorylation serves as the trigger for ATM activation. Perhaps the most substantive challenge to the Bakkenist and Kastan model relates to the mechanism by which ATM senses DSBs. The original model posited that the DNA damage signal is transmitted to ATM through structural changes in chromatin. The new results indicate that the MRN complex forms a bridge between ATM and the DSB site, and delivers an undefined signal (perhaps a conformational alteration) that triggers ATM autophosphorylation and monomer formation. In their studies, Bakkenist and Kastan found that treatment of cells with chromatin-disrupting agents, such as chloroquine or histone deacetylase inhibitors (HDACs), stimulated ATM in the absence of detectable DNA damage (*3*). However, these drugs may well induce abnormal DNA structures, such as hairpins, that can be recognized and processed by MRN (*11*). It will be interesting to learn whether ATM activation induced by chloroquine or HDACs is also dependent on the MRN complex. Finally, a more detailed understanding of the structural basis of the interaction between ATM and MRN is sure to provide new insights into the mechanism of ATM activation. Considering the pace of research in this field, the ATM activation model may prove to be as dynamic as the process of DNA damage signaling itself.



The MRN complex and ATM activation. (Step 1) The induction of DSBs in DNA leads to prompt recruitment of MR or MRN complexes. These complexes form a bridge between free DNA ends via the coiled-coil arms of Rad50 dimers. Inactive ATM dimers are recruited to the DSBs through interaction with the carboxyl terminus of Nbs1, and possibly a less stable interaction with Rad50. **(Step 2)** The 3′,5′-exonuclease activity of MRE11 catalyzes resection of free DNA ends, creating ssDNA. Activating signals are delivered to ATM dimers, possibly through a conformational change in Nbs1. ATM undergoes phosphorylation at Ser^{1981} accompanied by its conversion from a dimer to a monomer. The MR complex may also trigger a conformational change in ATM that stimulates substrate recruitment. **(Step 3)** Activated ATM monomers either remain in the vicinity of the DSB, where they phosphorylate colocalized substrates, or diffuse away from the DSB sites to phosphorylate nuclear substrates, such as p53 and Creb.

CREDIT: PRESTON HUEY/*SCIENCE*

References

1. Y. Shiloh, *Nat. Rev. Cancer* **3**, 155 (2003).
2. J.-H. Lee, T. T. Paull, *Science* **308**, 551 (2005); published online 24 March 2005 (10.1126/science.1108297).
3. C. J. Bakkenist, M. B. Kastan, *Nature* **421**, 499 (2003).
4. J. H. Petrini, *Curr. Opin. Cell Biol.* **12**, 293 (2000).
5. C. J. Bakkenist, M. B. Kastan, *Cell* **118**, 9 (2004).
6. J. Falck, J. Coates, S. P. Jackson, *Nature* **434**, 605 (2005).
7. V. Costanzo, T. Paull, M. Gottesman, J. Gautier, *PLoS Biol.* **2**, E110 (2004).
8. C. T. Carson *et al.*, *EMBO J.* **22**, 6610 (2003).
9. J. H. Lee, T. T. Paull, *Science* **304**, 93 (2004).
10. L. Zou, S. J. Elledge, *Science* **300**, 1542 (2003).
11. D. D'Amours, S. P. Jackson, *Nat. Rev. Mol. Cell Biol.* **3**, 317 (2002).

10.1126/science.1112069

Transduction of Receptor Signals by β-Arrestins

Robert J. Lefkowitz[1,2*] and Sudha K. Shenoy[2]

The transmission of extracellular signals to the interior of the cell is a function of plasma membrane receptors, of which the seven transmembrane receptor family is by far the largest and most versatile. Classically, these receptors stimulate heterotrimeric G proteins, which control rates of generation of diffusible second messengers and entry of ions at the plasma membrane. Recent evidence, however, indicates another previously unappreciated strategy used by the receptors to regulate intracellular signaling pathways. They direct the recruitment, activation, and scaffolding of cytoplasmic signaling complexes via two multifunctional adaptor and transducer molecules, β-arrestins 1 and 2. This mechanism regulates aspects of cell motility, chemotaxis, apoptosis, and likely other cellular functions through a rapidly expanding list of signaling pathways.

Transmission of extracellular signals across the plasma membrane by receptor-mediated signaling is one of the most fundamental cellular processes. However, only a small number of paradigmatic, generally operative mechanisms to accomplish this goal have been delineated. Examples include gating of ion channels, stimulation of heterotrimeric G proteins, and activation of receptor tyrosine kinases. However, recent findings have revealed another. A regulatory process, originally discovered as the means by which seven transmembrane receptor (7TMR) activation of G proteins is "desensitized" or turned off, quite surprisingly has been found to serve also as a parallel means of signal transduction. This newly appreciated signaling mechanism involves two families of proteins, the G protein–coupled receptor kinases (GRKs) and β-arrestins. The latter serve as multifunctional adaptor and scaffold proteins that recruit a broad spectrum of signaling molecules and assemblies to the receptors in a strictly activation-dependent fashion.

Historical Perspective and Classical Paradigms

In the mid-1980s, it was discovered that the prototypic adenylyl cyclase–coupled β_2-adrenergic receptor (β_2AR) for catecholamines and the visual sensing protein rhodopsin shared conserved structural and regulatory features, including sequence similarity, a seven transmembrane (TM) architecture, and a mechanism for "desensitization." This immediately suggested that all so-called G protein–coupled receptors (GPCRs) might be members of the same gene family and share these attributes—a hypothesis that was quickly confirmed [see (*1*) and references therein for a review of this historical material].

Today, we know that 7TMRs represent the largest (*2*), most versatile, and most ubiquitous of the several families of membrane receptors. Moreover, they are the most common target of therapeutic drugs (*3*). In response to a remarkable range of stimuli, including neurotransmitters, hormones, ions, and sensory stimuli, these receptors regulate the metabolism, secretory properties, electrical activity, shape, and motility of virtually all mammalian cells.

Studies of rhodopsin and the β_2AR also revealed that G protein–mediated signaling is attenuated or desensitized by a highly conserved process (*4*) that involves phosphorylation of the activated receptors by specific protein kinases, such as rhodopsin kinase (now known

Table 1. A list of β-arrestin–interacting proteins. ARF, ADP ribosylation factor; ARNO, ARF nucleotide exchange factor; IκBα, inhibitor of nuclear factor κB; PDE4D, phosphodiesterase 4D; PP2A, protein phosphatase 2A; Ral, members of the Ras superfamily of small guanosine triphosphatases (GTPases); Ral-GDS, Ral guanosine diphosphate (GDP) dissociation stimulator; RhoA, a small GTPase; small G/GEFs, small GTPase and guanine nucleotide exchange factors.

Binding protein	β-Arrestin isoform	Functional consequence	Ref.
		Trafficking proteins	
Clathrin	β-Arrestin 1, 2	Endocytosis	(*57*)
AP2	β-Arrestin 1, 2	Endocytosis	(*58*)
NSF	β-Arrestin 1	Endocytosis; recycling	(*59*)
		Small G/GEFs	
ARF6	β-Arrestin 2≫1	Endocytosis	(*60*)
ARNO	β-Arrestin 2	Endocytosis	(*60*)
Ral-GDS	β-Arrestin 1, 2	Ral-mediated cytoskeletal changes	(*61*)
RhoA	β-Arrestin 1	Angiotensin II-dependent stress fiber formation	(*62*)
		Signaling proteins	
MAPK cascade components			
ASK1	β-Arrestin 1, 2	JNK3 and p38 activation	(*24*)
c-Raf-1	β-Arrestin 1, 2	ERK activation	(*22, 23*)
JNK3	β-Arrestin 2≫>1	Stabilization of pJNK on endosomes	(*24*)
ERK2	β-Arrestin 1, 2	Stabilization of pERK on endosomes	(*22, 23, 26*)
Nonreceptor tyrosine kinases			
c-Src	β-Arrestin 1, 2	Endocytosis, ERK activation	(*15*)
Yes	β-Arrestin 1	Gαq activation and GLUT4 transport	(*18*)
Hck	β-Arrestin 1	Exocytosis of granules in neutrophils	(*17*)
Fgr	β-Arrestin 1	Exocytosis of granules in neutrophils	(*17*)
Others			
Mdm2	β-Arrestin 1, 2	Ubiquitination, endocytosis	(*10*)
IκBα	β-Arrestin 1, 2	Stabilization of IκBα upon β2AR and TNFR stimulation	(*63, 64*)
PDE4D family	β-Arrestin 1, 2	cAMP degradation	(*65*)
Dishevelled	β-Arrestin 1	Increase in TCF/LEF transcription	(*66*)
Dishevelled	β-Arrestin 2	Endocytosis of Frizzled4	(*42*)
PP2A	β-Arrestin 1	Ser^{412} dephosphorylation	(*49*)

[1]Howard Hughes Medical Institute, [2]Duke University Medical Center, Durham, NC 27710, USA.

*To whom correspondence should be addressed. E-mail: lefko001@receptor-biol.duke.edu

as GRK1) and the β-adrenergic receptor kinase or βARK (now known as GRK2), respectively. However, the receptor turn-off mechanism was found to require more than stimulus-dependent receptor or rhodopsin phosphorylation. In vitro, rhodopsin kinase–catalyzed phosphorylation of activated rhodopsin led to only partial quenching of rhodopsin signaling. A highly abundant and immunogenic retinal protein, now known as ''arrestin,'' was found to potentiate the signal-dampening effects of rhodopsin kinase phosphorylation on rhodopsin. Simultaneously, purification of βARK was found to lead to progressive loss of its ability to desensitize β_2ARs, in vitro. This suggested that some other factor necessary for the desensitization was being lost. Visual arrestin (also called arrestin 1), albeit at high concentrations, would restore this activity. Because expression of arrestin is limited to the retina, structural and functional homologs were hypothesized to exist in other tissues. Cloning of visual arrestin led to the identification of similar genes encoding β-arrestin 1 and β-arrestin 2 (also known as arrestin 2 and 3).

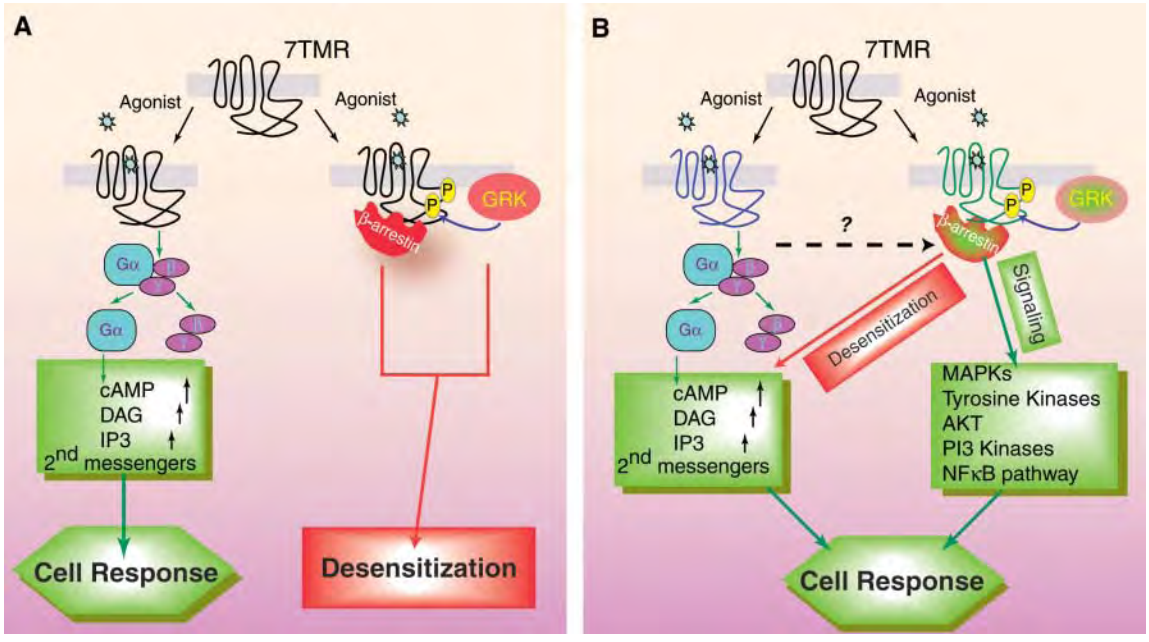


Fig. 1. Signal transduction by seven transmembrane receptors. (**A**) Classical paradigm. The active form of the receptor (R*) stimulates heterotrimeric G proteins and is rapidly phosphorylated by G protein–coupled receptor kinases (GRKs), which leads to β-arrestin recruitment. The receptor is thereby desensitized, and the signaling is stalled. (**B**) New paradigm. β-Arrestins not only mediate desensitization of G protein–signaling but also act as signal transducers themselves.

β-Arrestins 1 and 2 have marked specificity for binding phosphorylated β_2AR as opposed to phospho-rhodopsin, whereas the reverse is true of visual arrestin. Both agonist-induced conformational changes in the receptor and receptor phosphorylation contribute to driving receptor β-arrestin interaction, and β-arrestin competes with G_s for receptor interaction. This ''desensitization'' mechanism appears to be universal for 7TMRs (*4*), although some recent studies suggest that the visual arrestins do not strictly lead to signal termination but rather contribute to adaptation to varying light intensities (*5*).

The arrestins and GRKs are each members of small gene families (tables S1 and S2) (*6*, *7*). There are four arrestin genes and seven GRKs. Retinal rods and cones each have their own dedicated regulatory systems: Arrestin 1 and GRK 1 in the rods regulate rhodopsin; arrestin 4 (X arrestin) and GRK 7 in the cones regulate color opsins. β-Arrestins 1 and 2 and GRKs 2, 3, 5, and 6 are widely expressed and regulate most 7TMRs.

β-Arrestins Are Multifunctional Endocytic Adaptors and Signal Transducers

For signal transduction, 7TMRs propagate a chain of protein conformational changes in response to agonist stimulation. Thus, an essential characteristic of any general transducer of 7TMR signaling is its ability to interact universally with the receptors in an activation-dependent way and, thereby, to undergo conformational changes. Only three families of proteins have this attribute: heterotrimeric G proteins, β-arrestins, and GRKs. It is thus perhaps not surprising that recently β-arrestins have been found to mediate a variety of receptor signaling and regulatory processes and to bind to a growing list of endocytic and signaling proteins (Table 1 and table S3) (*8*).

Endocytosis. Early evidence that β-arrestins have larger roles in 7TMR biology than just desensitization came from the discovery that they also function as endocytic adaptors, linking receptors to the clathrin-coated pit machinery (*8*, *9*). Originally found to bind to clathrin itself, β-arrestins are now known to interact with other endocytic elements, including the adaptor protein AP2, the small guanosine triphosphatase ARF6 and its guanine nucleotide exchange factor ARNO, and the *N*-ethylmaleimide–sensitive fusion protein (NSF). β-Arrestins also bind and are ubiquitinated by the E3 ubiquitin ligase Mdm2 (*10*). This agonist-stimulated ubiquitination event is required for β-arrestin–mediated endocytosis, but precisely how or why is not known.

The avidity with which β-arrestins bind to activated phosphorylated receptors is largely determined by the pattern of GRK-mediated phosphorylation, generally on the C-terminal tails of the receptors (*11*). In the case of some receptors (''class A''), such as the β_2AR, β-arrestins bind relatively weakly, target the receptors to clathrin-coated pits, and then dissociate as the receptors internalize. For other receptors (''class B''), for example, the V_2R vasopressin receptor (V_2R) and the angiotensin II type 1a receptor (AT1aR), because of much tighter binding β-arrestin does not dissociate from the receptor and accompanies it into the cell, where the complex may reside for extended periods in endosomal vesicles before being sorted to lysosomes or slowly recycled. Patterns of 7TMR endocytosis parallel the kinetics of β-arrestin ubiquitination and de-ubiquitination: Transient β-arrestin ubiquitination correlates with class A and more persistent β-arrestin ubiquitination with class B behavior (fig. S1) (*12*). Endocytosis of the receptors plays roles in receptor resensitization by dephosphorylation, receptor recycling, receptor down-regulation, and receptor signaling (*9*).

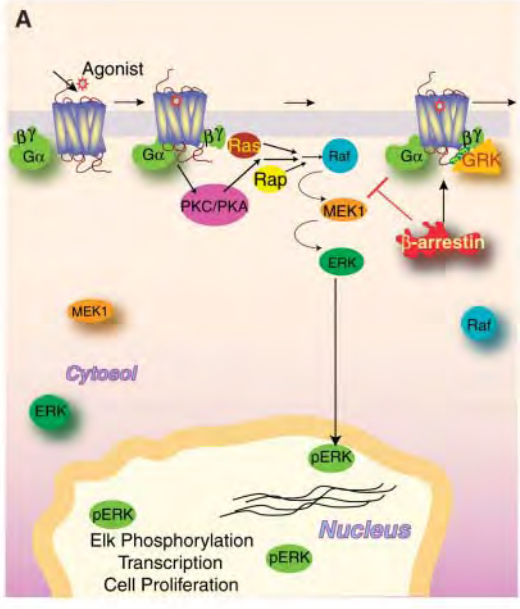
Signaling

The classical paradigms for signaling and desensitization of 7TMRs were developed within the context of providing a molecular

framework for understanding how receptor stimulation leads to regulation of the intracellular concentration of second messengers such as cyclic AMP (cAMP) (Fig. 1A). These events are essentially confined within the plasma membrane. However, the focus of attention has gradually shifted to a group of complex signaling pathways that ultimately link 7TM and other types of plasma membrane receptors to events in the cytoplasm and nucleus.

Initiation of such pathways generally requires that the stimulated receptors nucleate formation and activation of multicomponent signaling complexes and, in some cases, direct them to specific cellular destinations. Much evidence indicates that stimulus-dependent receptor recruitment of β-arrestins provides a general strategy used by 7TM, and perhaps other types of receptors, to accomplish these goals. β-Arrestins serve as adaptors, scaffolds, and/or signal transducers, and they connect the activated receptors with diverse signaling pathways within the cell (Fig. 1B, Table 1; table S3) (*8*, *13*).

Nonreceptor tyrosine kinases. Although 7TMRs and receptor tyrosine kinases were once viewed as completely distinct, nonoverlapping signaling mechanisms, it is now clear that some receptor tyrosine kinases activate G proteins, whereas some 7TMRs "transactivate" receptor tyrosine kinases, such as the epidermal growth factor (EGF) and platelet-derived growth factor (PDGF) receptors (*14*). Other 7TMRs co-opt the activity of nonreceptor tyrosine kinases, and in this process, β-arrestins have positive roles in signal transduction. At the β_2AR (*15*) and the Neurokinin 1 (NK1) receptor for substance P (*16*), β-arrestins mediate agonist-dependent recruitment of c-Src to the receptor, facilitating activation of the mitogen-activated protein (MAP) kinases extracellular signal–regulated kinases (ERK1 and ERK2). Other members of the c-Src family of nonreceptor tyrosine kinases such as Hck (*17*), Fgr (*17*), and Yes (*18*) are recruited to various 7TMRs by β-arrestins. This recruitment regulates such disparate physiological processes as granule release from human neutrophils (CXCR1) (*17*), antiapoptotic effects in response to NK1 receptors (*16*), and GLUT4 (a glucose transmembrane transporter) translocation in response to endothelin A receptors (*18*).



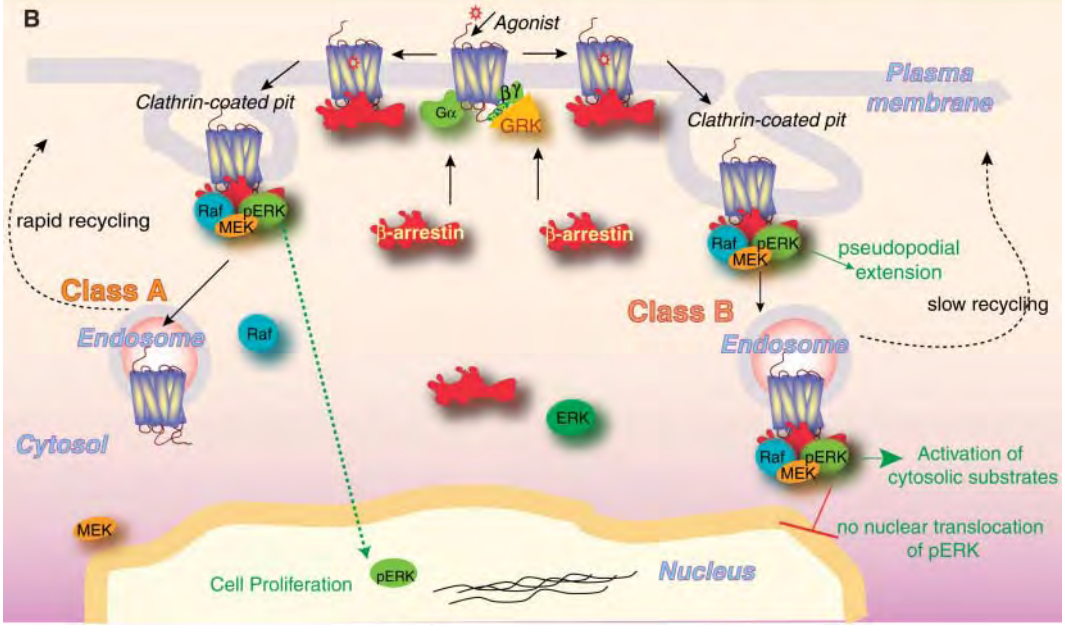


Fig. 2. 7TMR-stimulated pERK. (**A**) G protein–dependent ERK activation. Stimulation of G proteins activates the Raf family proteins through several convergent pathways. Raf translocates to plasma membrane on activation. (**B**) β-Arrestin–dependent ERK activation. β-Arrestin binding to phosphorylated receptors interdicts G protein–dependent signaling, while initiating new waves of signal transduction, for example, by activating ERK1/2.

The example of MAP kinases. A well-studied β-arrestin–dependent signaling system leads to activation of the MAP kinase ERK. MAP kinases are the terminal elements of highly conserved kinase cascades consisting of MAPKKKs (such as Raf), MAPKKs (such as MEK), and the MAP kinases themselves. There are three families of the multifunctional MAP kinases that include five ERKs, three c-Jun NH_2-terminal kinases (JNKs), and four p38s. Each kinase is activated by phosphorylation by the preceding kinase in the cascade. There are a dozen or more enzymes at each level (*19*).

The 7TMRs are connected to these MAPK signaling pathways by classical G protein–stimulated synthesis of second messengers (Fig. 2A) or by nonclassical pathways modulating novel effectors (*20*). MAP kinases activated in this way translocate from the cytosol to the nucleus, where they phosphorylate and activate transcription factors, which regulate programs of transcription that lead to proliferation, differentiation, and many other cellular processes (*19*). However, MAP kinases can also phosphorylate various cytosolic substrates, which leads to distinct but less well characterized consequences such as changes in cell shape and motility (*21*).

A conundrum has been how, in the face of the large number of enzymes in the different MAPK cascades—in which

a single upstream kinase can phosphorylate multiple downstream kinases—the cell is able to organize particular MAPK pathways with any specificity, reproducibility, or efficiency. The answer lies in ''scaffold'' proteins—molecules that tie together the appropriate kinases in a series (*19*). This accomplishes several goals: It forms a discrete signaling module, localizes the grouped components to specific areas in the cell, and insulates the active kinases from dephosphorylation by phosphatases. β-Arrestin 2 acts as a scaffold that binds all the component kinases of both the module that activates ERK1 and ERK2 (Raf, MEK, ERK) (*22*, *23*) and the module that activates JNK3 (ASK1, MKK4, JNK3) (*24*). Of the several mammalian MAP kinase scaffolds, β-arrestin 2 is the only one that is controlled by receptor stimulation.

β-Arrestin–mediated activation of ERK appears to be intimately linked to the function of β-arrestins in mediating endocytosis of receptors in clathrin-coated pits (Fig. 2B). The β-arrestin scaffolded signaling complex is internalized with class B receptors such as the AT1aR and ultimately is found in endocytic vesicles together with the receptors (*22*). Here, its activity persists for prolonged periods, perhaps because the phosphorylated ERK is protected from MAP kinase phosphatases (*21*, *25*). Class A receptors promote much less persistent β-arrestin–mediated activation of ERK (*26*). This apparently relates to the requirement for stable association of β-arrestins with the receptors for this activity to occur (Fig. 2B).

The AT1aR mediates disparate effects of angiotensin on vasoconstriction, smooth muscle cell motility and growth, and aldosterone secretion. Receptor stimulation activates ERK1 and ERK2 by either G protein (G_q)– or β-arrestin–mediated signaling pathways (*27*, *28*). Studies with mutant ligands and receptors that activate one pathway or the other, or with β-arrestin small interfering RNA (siRNA) or specific inhibitors of protein kinase C (PKC), have helped to delineate the characteristics of these two independent pathways (Fig. 3, A and B) (*25*). G protein–mediated activation is rapid and transient and is blocked by PKC inhibitors. It leads to nuclear translocation of the activated ERK, with consequent regulation of transcriptional programs and cellular proliferation. In contrast, β-arrestin–mediated activity is characterized by slower onset, greater persistence, retention of the activated ERK in cytosolic endocytic vesicles, and an absence or paucity of transcriptional regulation. Such distinct characteristics for the ERK activated by these different mechanisms strongly imply distinct physiological consequences. For the β-arrestin pathway, these are likely to include effects on cell motility, chemotaxis, and apoptosis (see below).

For the AT1aR, ERK activation transduced through β-arrestin is mediated almost exclusively by β-arrestin 2, with physiological amounts of β-arrestin 1 serving primarily as an inhibitor (*29*). For the protease-activated receptor 2 (PAR2) (*21*) and NK1 receptors (*16*), however, β-arrestin 1 appears to promote activation of ERK, whereas, for the CCR7 receptor, β-arrestin 2 is implicated (*30*). AT1aR-mediated activation of ERK features independent G protein– and β-arrestin 2–mediated pathways. However, in other cases, G proteins and β-arrestins may act in sequence or in a concerted fashion. For example, with several chemokine receptors, ERK activation and chemotaxis are sensitive to both pertussis toxin (implicating signaling through G protein G_i) and siRNA-mediated depletion of β-arrestin 2 (*30*–*32*).

An interesting implication of the existence of β-arrestin (versus G protein)–mediated signaling is the potential for ligands to selectively activate one or the other pathway. For example, a mutated angiotensin peptide (SII angiotensin) activates β-arrestin–mediated, but not G protein–mediated, signaling to ERK (*25*, *27*). A physiological example of such pathway-selective ligands is provided by the two endogenous ligands, Epstein-Barr virus–induced receptor ligand chemokine (ELC, CCL19) and secondary lymphoid tissue chemokine (SLC, CCL21), of the chemokine receptor CCR7, which regulates the homing and trafficking of T lymphocytes. The ligands have equivalent abilities to activate G proteins, but only ELC leads to receptor phosphorylation, β-arrestin recruitment, and β-arrestin–dependent ERK activation (*30*). The differences in the portfolios of physiological responses evoked by these two different ligands remain to be determined.

Chemotaxis and cell motility. Chemotaxis, the directed migration of cells along a gradient of chemoattractant, relies on signals from 7TMRs for chemokines. The signaling pathways used are heterogeneous and quite complex. Although G proteins (most often G_i) are involved, β-arrestins have recently emerged as important transducers of some of these signals, perhaps through their ability to mediate activation of MAP kinases. Lymphocytes from knockout mice lacking β-arrestin 2, but not from mice lacking β-arrestin 1, are impaired in chemotactic responses to stromal cell–derived factor–1 (SDF-1) mediated by the CXCR4

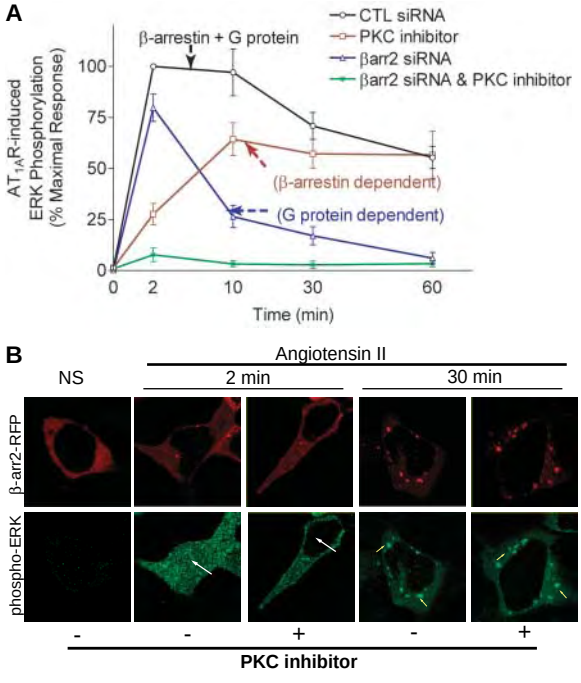


Fig. 3. 7TMR stimulation of ERK by distinct G protein– and β-arrestin–dependent pathways. (**A**) Temporal patterns. Angiotensin II stimulates ERK1 and ERK2 phosphorylation in HEK 293. The black activity curve (with circles) represents both Gαq/11- and β-arrestin–dependent pERK. After transfection of cells with β-arrestin2 siRNA, only a rapid and transient pool of angiotensin II–stimulated pERK is observed; as seen in the G-protein–dependent ERK activity curve (purple, with triangles). This Gαq/11–dependent pathway requires PKC activation, and hence, a PKC inhibitor such as Ro-31-8425 inhibits this mechanism of ERK activation. The red line (with squares) represents β-arrestin–mediated ERK signaling. The combination of β-arrestin 2 siRNA transfection and treatment with PKC inhibitor virtually eliminates angiotensin II–stimulated pERK (as seen in the green line with asterisk). Data taken from (*25*). (**B**) Subcellular distribution. Shown are confocal images of fixed HEK 293 cells expressing the AT1aR (not displayed) and β-arrestin 2-RFP (red). Cellular distribution of pERK was visualized by immunolabeling with a polyclonal antibody against phospho-ERK1/2 followed by a Bodipy fluorescein-conjugated secondary antibody (green). In nonstimulated cells, very little or no pERK is detected by immunostaining (green, bottom row, leftmost panel). The second column of confocal panels displays the pERK detectable in the cytoplasm and nucleus after 2 min of angiotensin II treatment in the absence of the PKC inhibitor (Ro-31-8425). The third column shows that in cells pretreated with the PKC inhibitor, there is marked inhibition of the G protein–stimulated 2-min pERK signal. pERK in the nucleus is completely abolished, whereas some pERK is still detectable in the cytoplasm and at the plasma membrane. The fourth and fifth columns show cells stimulated for 30 min with angiotensin II. At this time point, only β-arrestin–dependent pERK activity persists and is exclusively present on endocytic vesicles. This pERK signal is completely insensitive to the PKC inhibitor, which indicates that it is independent of the G protein pathway. [Data are condensed and reproduced from (*25*) with permission from the *Journal of Biological Chemistry*.]

receptor (*32*). The same is true of human embryonic kidney (HEK) 293 cells transfected with CXCR4 in which expression of β-arrestin 2 has been lowered by siRNA. In these cells, deficiency of β-arrestin 2 also impairs activation of p38 MAP kinase by SDF1, and inhibitors of this kinase, but not of ERK1 and ERK2, block chemotaxis. Thus, in the HEK 293 cells, β-arrestins may mediate the SDF-1 chemotactic signal by activation of p38 (*31*).

Chemotactic responses to PAR2, a 7TMR for trypsin and related proteases that is abundant in highly motile cells such as neutrophils, macrophages, and tumor cells, also require signaling through β-arrestins (*21*). In NIH 3T3 cells transfected with PAR2, the signaling involves ERK1 and ERK2 and β-arrestin–dependent reorganization of the actin cytoskeleton. Increased β-arrestin–mediated assembly of PAR2 and ERK at the leading edge of the cells in the extending pseudopodia leads to prolonged ERK activation (*21*). β-Arrestin–nucleated assembly of activated ERK in pseudopodia is also operative in the highly metastatic human MDA MB-231 breast cancer cell line. These cells, in contrast with the much less metastatic MDA MB-468 line, secrete a trypsin-like PAR2 ligand, which by an autocrine mechanism stimulates cell migration. This process requires both β-arrestin 1 and β-arrestin 2, which may have distinct roles in contributing to the migratory behavior of the metastatic cells by β-arrestin–dependent ERK activation or other mechanisms (*33*).

Inhibition of apoptosis. β-Arrestin contributes to antiapoptotic signaling. In the case of the NK1 receptor for substance P, this may be mediated by ERK activation (*16*). For the insulin-like growth factor 1 (IGF-1) receptor, it is mediated through a β-arrestin 1–dependent pathway leading to activation of phosphatidylinositol 3-kinase (PI3K) and AKT, which operates independently of the tyrosine kinase activity of the IGF-1 receptor (*34*). In mouse embryonic fibroblasts, stimulation of various 7TMRs causes apoptosis in the absence of, but not in the presence of, β-arrestins, which again suggests a role for the β-arrestins in antiapoptotic signaling (*35*).

Conformational changes in β-arrestins. The structure of β-arrestin 2, as modeled on the atomic structure of β-arrestin 1 in its basal state (*36*), reveals an elongated molecule ~70 Å. It has two distinct domains linked by a 12-residue "hinge" region. The domains are held intact by intramolecular interactions defined by buried polar residues, as well as the "three-element interface" that comprises the "buried" C tail, β strand I, and α helix I. β-Arrestins interact with many different protein partners (Fig. 4, A and B), yet they do not have any well-characterized domains specialized for protein-protein interactions. β-Arrestins act as phosphoprotein sensors, which generally display higher affinity for the phosphorylated forms of their binding partners.

Interaction of β-arrestins with activated receptors induces global conformational changes and rapid posttranslational modifications (dephosphorylation and ubiquitination) of the β-arrestin molecule (Fig. 4C) (*37*). Disruption of the polar core by phosphate moieties on receptors and the resulting rearrangement of the "three-element interface" is proposed to induce activating conformational changes in the β-arrestins. Such 7TMR activation of β-arrestin enhances its interaction with several of its binding partners. For example, in some cases, receptor stimulation instigates β-arrestin binding to signaling partners, such as c-Src and phosphorylated ERK1 or ERK2, or to endocytic partners such as clathrin or AP2 (*8*). This reflects the conformational changes in β-arrestin that occur on binding to the phosphorylated domains of activated 7TMRs (Fig. 4C). Moreover, it is possible that multiple conformational states of β-arrestins exist, induced by binding to receptors phosphorylated on different sites or different combinations of sites, which then lead to distinct functional outcomes. Conversely, β-arrestins stabilize conformational changes in the receptors. Thus, in analogy with the high-affinity ternary complex of agonist, receptor, and G protein (*38*), high-affinity ternary complexes of agonist, receptor, and β-arrestin are also formed (*39*).

Variations on a theme. In addition to GRK-dependent interaction of β-arrestins with conventional 7TMRs, arrestins also interact with other receptor types. In mammalian cells, GRK2 and β-arrestin 2 interact with the developmentally important 7TM molecule smooth-

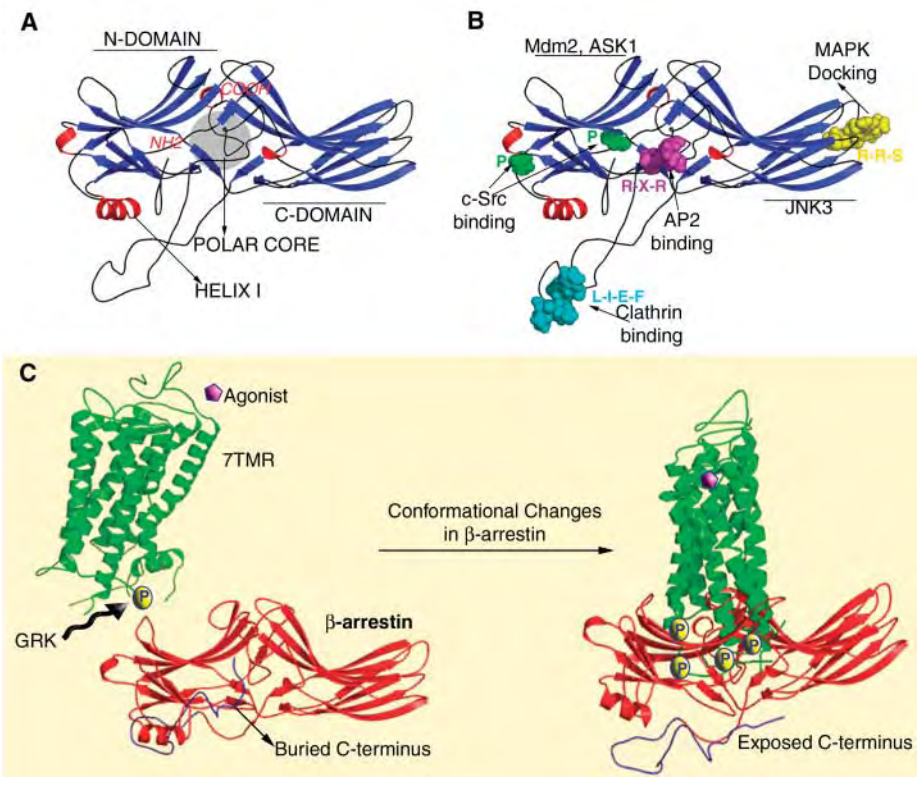


Fig. 4. β-Arrestin 2: Structural model and receptor-dependent conformational changes. (**A**) Basic structural attributes of β-arrestin. Portrayed is a structural model of rat β-arrestin 2 derived by homology modeling using the available structures of arrestin and β-arrestin 1 (PDB files: 1G4R, 1G4M, 1CF1, and 1JSY) and compiled with the program *PyMOL* (*56*). β Sheets are colored blue, helix I (indicated) is red, and the connecting loops and C-tail are black. N-domain and C-domain regions are connected by a hinge region. Both the N-terminal β strand I and the C-terminal β strand XX are juxtaposed to helix I. (**B**) Protein-protein interaction motifs mapped on β-arrestin 2. Prolines at positions 91 and 121 are important for β-arrestin 1–cSrc interaction and are depicted as green space filling spheres. The clathrin-binding domain L-I-E-F and the MAPK docking domain R-R-S are shown as cyan and yellow spheres, respectively. AP2 binding requires the arginine residues at positions 394 and 396, shown as magenta spheres. Mdm2 and ASK1 bind to regions delimited by residues, 160 to 300 and 1 to 185 (N-terminal half), respectively. JNK3 binds to the region 185 to 410 (C-terminal half). Both N- and C-domains of β-arrestin interact with 7TMRs. (**C**) 7TMR-induced conformational changes in β-arrestin. Structural models of 7TMR (rhodopsin: source, PDB file 1L9H) and rat β-arrestin 2 are shown in the basal state on the left side. The C-terminal tail of β-arrestin (blue) is buried and not accessible. Agonist stimulation leads to phosphorylation of serine and threonine residues on the receptor C tail. The charged domain thus created penetrates and disrupts the polar core of β-arrestin and leads to global conformational rearrangements, which cause the β-arrestin C tail to be released and exposed for protein interactions. The C terminus of β-arrestin contains the clathrin- and AP2-binding regions [see (B)].

ened (Smo) in an activity-dependent manner and regulate its endocytosis (*40*). In zebrafish embryos, knockdown of β-arrestin 2 by morpholino antisense leads to a lethal phenotype remarkably similar to that observed after genetic knockout of Smo or its downstream effector Gli (*41*).

Wnts are important morphogenetic and developmental ligands whose actions are mediated by the 7TM frizzled receptors. β-Arrestin 2 mediates endocytosis of the Wnt 5A-Fz-4 complex, but recruitment of β-arrestin is accomplished by an intermediary adaptor, disheveled, which interacts with β-arrestin 2 in a PKC-dependent fashion (*42*).

Even further afield are receptors that are not members of the 7TM superfamily. The tyrosine kinase IGF-1 receptor not only activates G proteins (*43*) but also recruits β-arrestin 1 after IGF-1 stimulation (*44*, *45*). This mediates activation of PI3K (*34*), AKT (*34*), and ERK1/2 (*44*). The transforming growth factor–β (TGF-β) family of ligands signal through heteromeric complexes formed by members of three single membrane–spanning receptor families (TβRI to III). β-Arrestin 2 mediates endocytosis of TβRIII, after it is phosphorylated on threonine (Thr^{841}) by TβRII, which is itself a serine kinase (*46*). This down-regulates antiproliferative signaling.

Future Directions

Research in cellular signal transduction is evolving from a focus on linear pathways to a broader view of signaling networks composed of interacting pathways (*47*). In the terminology of this field, β-arrestins serve both as nodes, which receive signals from multiple receptor inputs, and junctions, which route signals to various effectors. Signaling cross-talk can result from these functions. For example, stimulation of the β_2AR or the insulin receptor leads to activation of ERK, which phosphorylates β-arrestin 1 on Ser^{412} (*48*, *49*), a modification that impairs internalization of several 7TMRs and their consequent activation of ERKs (*49*, *50*).

For both the G_q-coupled AT1aR and the G_s-coupled V_2R, β-arrestin–mediated signaling to ERK requires prior receptor phosphorylation by GRKs 5 and 6 (*51*, *52*). Phosphorylation of the receptors by GRK2, while empowering receptor endocytosis or desensitization, actually antagonizes β-arrestin–mediated activation of ERK. It seems possible that the different GRKs phosphorylate distinct sets of sites on the carboxyl termini of the receptors and thus establish a ''bar code'' that somehow determines the conformation and, hence, the functional potential of the receptor-bound β-arrestin. It will be important to determine the applicability of these principles to other receptors and to other β-arrestin–mediated signaling phenomena.

A provocative issue is the potential role of β-arrestin ubiquitination in regulating its signaling functions. Might it serve as a recognition motif for assembling signaling complexes in analogy with its role in assembling elements of the protein-trafficking machinery? Several considerations are consistent with this hypothesis: the activation (receptor)–dependent nature of the modification (*10*), the correlation of the β-arrestin–mediated endocytic behavior of receptors (class A versus class B) with the stability of the ubiquitin modification (*12*), the correlation of the endocytic behavior of 7TMRs with the robustness of their β-arrestin–mediated activation of ERK (*26*), and the control of several signaling systems by ubiquitination (*53*).

The discovery of β-arrestin–mediated signaling highlights an emerging concept, that of ligand-directed signaling (*54*). In the simplest classical models, receptors exist in two states, active and inactive, with agonists stabilizing the active state, thereby driving activation of effectors such as G proteins. However, the conformation of a receptor that interacts with G protein can be distinct from that which interacts with β-arrestins. Thus, for both the β_2AR and V_2R, inverse agonists for G protein signaling (that is, compounds that actually lower basal adenylyl cyclase) are stimulatory agonists for β-arrestin–mediated signaling (*55*). The ability of ligands to differentially favor one or the other conformation suggests significantly greater diversity and fine-tuning of signaling possibilities for a single receptor than previously imagined. Moreover, such putative β-arrestin– or G protein–specific ligands might have valuable therapeutic properties and perhaps more restricted side effects.

References and Notes

1. R. J. Lefkowitz, *Trends Pharmacol. Sci.* **25**, 413 (2004).
2. R. Fredriksson, M. C. Lagerstrom, L. G. Lundin, H. B. Schioth, *Mol. Pharmacol.* **63**, 1256 (2003).
3. T. Gudermann, B. Nurnberg, G. Schultz, *J. Mol. Med.* **73**, 51 (1995).
4. T. A. Kohout, R. J. Lefkowitz, *Mol. Pharmacol.* **63**, 9 (2003).
5. V. Y. Arshavsky, *Sci. STKE* **2003**, pe43 (2003).
6. J. A. Pitcher, N. J. Freedman, R. J. Lefkowitz, *Annu. Rev. Biochem.* **67**, 653 (1998).
7. J. G. Krupnick, J. L. Benovic, *Annu. Rev. Pharmacol. Toxicol.* **38**, 289 (1998).
8. S. K. Shenoy, R. J. Lefkowitz, *Biochem. J.* **375**, 503 (2003).
9. A. Claing, S. A. Laporte, M. G. Caron, R. J. Lefkowitz, *Prog. Neurobiol.* **66**, 61 (2002).
10. S. K. Shenoy, P. H. McDonald, T. A. Kohout, R. J. Lefkowitz, *Science* **294**, 1307 (2001).
11. R. H. Oakley, S. A. Laporte, J. A. Holt, M. G. Caron, L. S. Barak, *J. Biol. Chem.* **275**, 17201 (2000).
12. S. K. Shenoy, R. J. Lefkowitz, *J. Biol. Chem.* **278**, 14498 (2003).
13. http://stke.sciencemag.org/cgi/cm/stkecm;CMP_15654, for subscribers.
14. N. J. Pyne, C. Waters, N. A. Moughal, B. S. Sambi, S. Pyne, *Biochem. Soc. Trans.* **31**, 1220 (2003).
15. L. M. Luttrell *et al.*, *Science* **283**, 655 (1999).
16. K. A. DeFea *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11086 (2000).
17. J. Barlic *et al.*, *Nat. Immunol.* **1**, 227 (2000).
18. T. Imamura *et al.*, *J. Biol. Chem.* **276**, 43663 (2001).
19. D. K. Morrison, R. J. Davis, *Annu. Rev. Cell Dev. Biol.* **19**, 91 (2003).
20. K. L. Pierce, R. T. Premont, R. J. Lefkowitz, *Nat. Rev. Mol. Cell Biol.* **3**, 639 (2002).
21. L. Ge, Y. Ly, M. Hollenberg, K. DeFea, *J. Biol. Chem.* **278**, 34418 (2003).
22. L. M. Luttrell *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2449 (2001).
23. K. A. DeFea *et al.*, *J. Cell Biol.* **148**, 1267 (2000).
24. P. H. McDonald *et al.*, *Science* **290**, 1574 (2000).
25. S. Ahn, S. K. Shenoy, H. Wei, R. J. Lefkowitz, *J. Biol. Chem.* **279**, 35518 (2004).
26. A. Tohgo *et al.*, *J. Biol. Chem.* **278**, 6258 (2003).
27. H. Wei *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 10782 (2003).
28. http://stke.sciencemag.org/cgi/cm/stkecm;CMP_16109, for subscribers.
29. S. Ahn, H. Wei, T. R. Garrison, R. J. Lefkowitz, *J. Biol. Chem.* **279**, 7807 (2004).
30. T. A. Kohout *et al.*, *J. Biol. Chem.* **279**, 23214 (2004).
31. Y. Sun, Z. Cheng, L. Ma, G. Pei, *J. Biol. Chem.* **277**, 49212 (2002).
32. A. M. Fong *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7478 (2002).
33. L. Ge, S. K. Shenoy, R. J. Lefkowitz, K. DeFea, *J. Biol. Chem.* **279**, 55419 (2004).
34. T. J. Povsic, T. A. Kohout, R. J. Lefkowitz, *J. Biol. Chem.* **278**, 51334 (2003).
35. C. M. Revankar, C. M. Vines, D. F. Cimino, E. R. Prossnitz, *J. Biol. Chem.* **279**, 24578 (2004).
36. M. Han, V. V. Gurevich, S. A. Vishnivetskiy, P. B. Sigler, C. Schubert, *Structure* (*Camb.*) **9**, 869 (2001).
37. V. V. Gurevich, E. V. Gurevich, *Trends Pharmacol. Sci.* **25**, 105 (2004).
38. A. De Lean, J. M. Stadel, R. J. Lefkowitz, *J. Biol. Chem.* **255**, 7108 (1980).
39. V. V. Gurevich, R. Pals-Rylaarsdam, J. L. Benovic, M. M. Hosey, J. J. Onorato, *J. Biol. Chem.* **272**, 28849 (1997).
40. W. Chen *et al.*, *Science* **306**, 2257 (2004).
41. A. M. Wilbanks *et al.*, *Science* **306**, 2264 (2004).
42. W. Chen *et al.*, *Science* **301**, 1391 (2003).
43. L. M. Luttrell *et al.*, *J. Biol. Chem.* **270**, 16495 (1995).
44. F. T. Lin, Y. Daaka, R. J. Lefkowitz, *J. Biol. Chem.* **273**, 31640 (1998).
45. http://stke.sciencemag.org/cgi/cm/stkecm;CMP_15950, for subscribers.
46. W. Chen *et al.*, *Science* **301**, 1394 (2003).
47. J. D. Jordan, E. M. Landau, R. Iyengar, *Cell* **103**, 193 (2000).
48. F. T. Lin *et al.*, *J. Biol. Chem.* **272**, 31051 (1997).
49. C. J. Hupfeld, J. L. Resnik, S. Ugi, J. M. Olefsky, *J. Biol. Chem.* **280**, 1016 (2005).
50. F. T. Lin, W. E. Miller, L. M. Luttrell, R. J. Lefkowitz, *J. Biol. Chem.* **274**, 15971 (1999).
51. J. Kim *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 1442 (2005).
52. X. R. Ren *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 1448 (2005).
53. C. Wang *et al.*, *Nature* **412**, 346 (2001).
54. T. Kenakin, *Trends Pharmacol. Sci.* **24**, 346 (2003).
55. M. Azzi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 11406 (2003).
56. *The PyMOL Molecular Graphics System* (DeLano Scientific, San Carlos, CA, 2002), available at www.pymol.org.
57. O. B. Goodman Jr. *et al.*, *Nature* **383**, 447 (1996).
58. S. A. Laporte *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3712 (1999).
59. P. H. McDonald *et al.*, *J. Biol. Chem.* **274**, 10677 (1999).
60. A. Claing *et al.*, *J. Biol. Chem.* **276**, 42509 (2001).
61. M. Bhattacharya *et al.*, *Nat. Cell Biol.* **4**, 547 (2002).
62. W. G. Barnes *et al.*, *J. Biol. Chem.* **280**, 8041 (2005).
63. H. Gao *et al.*, *Mol. Cell* **14**, 303 (2004).
64. D. S. Witherow, T. R. Garrison, W. E. Miller, R. J. Lefkowitz, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8603 (2004).
65. S. J. Perry *et al.*, *Science* **298**, 834 (2002).
66. W. Chen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14889 (2001).

67. We thank E. Hall and D. Addison for excellent secretarial assistance. R.J.L. is a HHMI investigator. NIH Grants HL 16037 and HL 70631 (to R.J.L.) supported this work.

Supporting Online Material

www.sciencemag.org/cgi/content/full/308/5721/512/DC1
Materials and Methods
SOM Text
Figs. S1 and S2
Tables S1 to S3
References and Notes

10.1126/science.1109237

H_2S Induces a Suspended Animation–Like State in Mice

Eric Blackstone,[1,2] Mike Morrison,[2] Mark B. Roth[2]*

Many organisms respond to changes in environmental conditions by entering into a suspended animation–like state in which a decrease in metabolic rate (MR) is followed by a reduction in core body temperature (CBT) (*1*). Regulated induction of a hypometabolic state is hypothesized to have great medical benefit for a variety of conditions, including ischemia and reperfusion injury, pyrexia, and other trauma (*2*). Suspended animation–like states may also be useful for creating beneficial hypothermia in surgical situations and for improving organ preservation (*1*).

Inhibiting oxidative phosphorylation reversibly induces states of profound hypometabolism in several model organisms (*3*–*5*). Because hydrogen sulfide (H_2S) is a specific, potent, and reversible inhibitor of complex IV (cytochrome c oxidase), the terminal enzyme complex in the electron transport chain (*6*), we hypothesized that it could reduce MR and CBT in mammals.

When mice were exposed to 80 ppm of H_2S, their oxygen (O_2) consumption dropped by ~50% and their carbon dioxide (CO_2) output dropped by ~60% within the first 5 minutes (Fig. 1A) (*7*). If left in this environment for 6 hours, their MR dropped by ~90% (Fig. 1A). The MR of control mice, as judged from O_2 consumption and CO_2 output increases (*8*). This drop in MR was followed by a drop in CBT to ~2°C above ambient temperature (Fig. 1B). The average CBT of these mice reached a minimum of 15°C in an ambient temperature of 13°C (Fig. 1B). At this minimum CBT, both CO_2 output and O_2 consumption was ~10% of normal (Fig. 1A), and the breathing rate of the mice decreased from ~120 breaths per minute (BPM) to less than 10 BPM (*8*). After 6 hours of exposure to H_2S, the mice were returned to room air and temperature, and their MR and CBT returned to normal (Fig. 1, A and B).

Exposing mice to varying concentrations of H_2S revealed a linear relationship between the concentration of H_2S and CBT (Fig. 1C). CBT dropped faster and reached lower temperatures as concentrations of H_2S increased from 0 to 80 ppm (*8*), suggesting that the effects of H_2S are concentration-dependent. However, this MR reduction is not dependent on ambient temperature (fig. S1).

Because H_2S can be toxic in high doses, we conducted behavioral and functional tests, selected from the SHIRPA protocol (*9*), to assay for H_2S-induced damage. No behavioral or functional differences in the mice were detected after exposure to 80 ppm of H_2S for 6 hours (*8*). In the absence of H_2S, no effect on CBT was observed (Fig. 1B, control atmosphere). In addition, others report no long-term health effects with these H_2S concentrations (*6*).

The sequential drop in MR and CBT observed in mice (Fig. 1D) exposed to 80 ppm of H_2S is similar to that observed when animals initiate hibernation, daily torpor, or estivation (*1*). On-demand induction of a suspended animation–like state could provide insight into the mechanisms that govern natural states of reduced metabolism. Lowering metabolic demand in this way could be used to reduce physiological damage resulting from trauma and might improve outcomes after surgery.

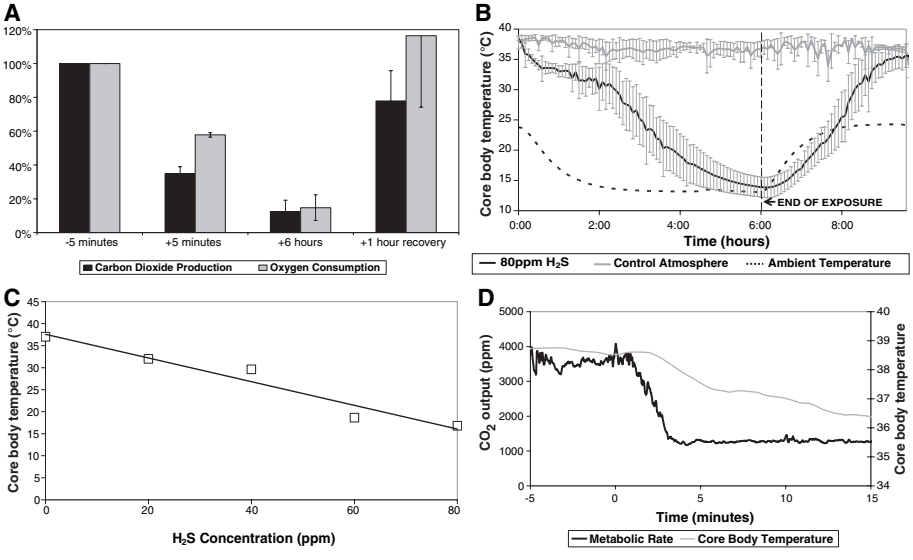


Fig. 1. CBT and MR of mice exposed to H_2S. (**A**) Relative CO_2 production and O_2 consumption of mice exposed to 80 ppm of H_2S. (**B**) CBT of mice during 6 hours of exposure to either 80 ppm of H_2S (black line) or the control atmosphere (gray line). The dotted line indicates ambient temperature. Values in (A) and (B) are means ± one standard deviation. (**C**) Linear relationship between H_2S concentration and CBT ($R^2 = 0.95$) after 6 hours of exposure. (**D**) CO_2 output and CBT of mice (time = 0 at the start of H_2S exposure).

References and Notes

1. H. V. Carey, M. T. Andrews, S. L. Martin, *Physiol. Rev.* **83**, 1153 (2003).
2. K. L. Drew, M. E. Rice, T. B. Kuhn, M. A. Smith, *Free Radical. Biol. Med.* **31**, 563 (2001).
3. P. A. Padilla, M. B. Roth, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7331 (2001).
4. T. G. Nystul, M. B. Roth, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9133 (2004).
5. K. B. Storey, J. M. Storey, *Q. Rev. Biol.* **65**, 145 (1990).
6. R. O. Beauchamp Jr. *et al.*, *Crit. Rev. Toxicol.* **13**, 25 (1984).
7. Materials and methods are available as supporting online material on *Science* Online.
8. E. Blackstone, M. Morrison, M. B. Roth, unpublished data.
9. D. C. Rogers *et al.*, *Mamm. Genome* **8**, 711 (1997).
10. We thank J. Goldmark, D. Miller, K. Chan, and A. J. Katzaroff for helpful discussion and commentary, J. Blackwood for instruction on the installation of telemetry devices, and Byrne Specialty Gases for gas science consultations. Supported by NIH grant no. GM48435 (M.R.), by donations to the Fred Hutchinson Cancer Research Center, and by National Institute of Aging award no. T32 AG00057 (E.B.).

Supporting Online Material
www.sciencemag.org/cgi/content/full/308/5721/518/DC1
Materials and Methods
SOM Text
Fig. S1

10 December 2004; accepted 10 March 2005
10.1126/science.1108581

[1]Molecular and Cellular Biology Program, University of Washington, Seattle, WA 98195, USA. [2]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98105, USA.

*To whom correspondence should be addressed. E-mail: mroth@fhcrc.org

Solar Wind Origin in Coronal Funnels

Chuan-Yi Tu,[1,3]* Cheng Zhou,[2] Eckart Marsch,[3] Li-Dong Xia,[4] Liang Zhao,[1] Jing-Xiu Wang,[5] Klaus Wilhelm[3]

The origin of the solar wind in solar coronal holes has long been unclear. We establish that the solar wind starts flowing out of the corona at heights above the photosphere between 5 megameters and 20 megameters in magnetic funnels. This result is obtained by a correlation of the Doppler-velocity and radiance maps of spectral lines emitted by various ions with the force-free magnetic field as extrapolated from photospheric magnetograms to different altitudes. Specifically, we find that Ne^{7+} ions mostly radiate around 20 megameters, where they have outflow speeds of about 10 kilometers per second, whereas C^{3+} ions with no average flow speed mainly radiate around 5 megameters. Based on these results, a model for understanding the solar wind origin is suggested.

The solar wind has been an essential topic in space physics ever since 1951 when it was inferred from cometary observations (*1*) and predicted from coronal models (*2*). Its properties were then analyzed through in situ observations (*3*, *4*). However, even after five decades of spacecraft exploration and theoretical modeling, the exact source of the solar wind in the solar atmosphere remains a mystery.

In the 1960s and 1970s, the "coronal base" (*5*–*10*) was assumed to be the inner boundary of the solar wind, yet its location and properties remained poorly defined. Considering that heating of the corona and acceleration of the solar wind are closely related, the models were extended throughout the solar transition region (TR) down to the chromosphere (*11*–*13*). It was suggested that the place where hydrogen becomes ionized might be considered as the inner boundary (*14*). In the past decade, models were developed with the assumption that the solar wind is created in coronal funnels (*15*–*18*), which are magnetic structures expanding into the corona with a narrow neck in the photosphere (*19*, *20*).

For observations and diagnostics of the solar wind source, the spectrometer SUMER (Solar Ultraviolet Measurements of Emitted Radiation) on SOHO (Solar and Heliospheric Observatory) provides a variety of spectral lines emitted in the chromosphere, transition region, or corona (*21*, *22*). However, solar disk observations resolved only two-dimensional (2-D) structures seen in planar projection and provided no height information along the line of sight (LOS). Hassler *et al.* (*23*) and Wilhelm *et al.* (*24*) identified the source regions of the fast solar wind by means of radiance and Doppler-shift maps in the polar coronal hole (CH), and Xia *et al.* (*25*, *26*) in equatorial CHs. The key tool for detecting outflow was the Doppler blueshift (indicating negative LOS velocities of the plasma away from the solar surface) of the emission line of Ne^{7+} at 77 nm. In these studies, the Ne VIII emission was compared with the Si II radiance in projected 2-D maps. However, the altitude of the emission regions remained unclear, as did the location where the solar wind acceleration actually starts.

Limb observations can reveal the height variation of the emission, but even then different structures are mixed together along the LOS. Moreover, they are perpendicular to the radial direction and therefore do not allow one to study radial outflow with the help of line-shift measurements. Although the coronal magnetic field cannot be measured directly, it can be constructed by various extrapolation methods (*27*) and so may provide some height information.

Only recently have magnetic fields, measured as photospheric magnetograms by MDI (Michelson Doppler Imager) (*28*) or Earth-based solar observatories, been used to determine the 3-D coronal field at low altitudes and spatial scales down to 1 Mm. The fields in active regions (AR) and CHs were constructed by force-free extrapolation from MDI data (*29*). The extrapolation technique is applied in a finite rectangular segment of the solar atmosphere and provides the coronal magnetic field $\mathbf{B}(x, y, z)$, with Cartesian components B_x, B_y, and B_z, where z is the vertical component, and x and y are the horizontal components, of the radial distance vector $\mathbf{r}$.

The model field has not yet been correlated at various heights with the ultraviolet emission patterns, especially in CHs where no bright loops occur. Such correlations are presented here to answer two important questions: At which height in CHs is the coronal base located, and where in coronal funnels does the solar wind flow start? We obtain the formation heights of spectral lines from a correlation of their radiances and Doppler maps with the extrapolated magnetic field and find that the acceleration starts between 5 Mm and 20 Mm above the photosphere and that speeds of about 10 km/s may be reached at 20 Mm.

Therefore, the inner boundary of the solar wind is neither at the old coronal base nor at the top of the chromosphere but should be defined as the point in the funnel where downflow changes into outflow, implying also sideflows consistent with supergranular convection. These results contrast with conventional (1-D) solar wind models (*30*), in which plasma starts flowing from about 2 Mm, whereby it is assumed that mass, momentum, and energy are conserved along the magnetic flux tube linking the upper chromosphere with the lower corona. The results obtained here shed new light on the magnetic structure and location of the solar wind source.

Correlations between radiation and magnetic field. We use data from the emission lines Si II (153.3 nm), C IV (154.8 nm), and Ne VIII (77.0 nm) measured by SUMER. These ultraviolet lines are emitted by the following ions: (i) Si^+, formed at an electron temperature of about 2×10^4 K in the chromosphere; (ii) C^{3+}, formed at 1×10^5 K in the TR, and (iii) Ne^{7+}, formed at 6×10^5 K in the lower corona. The observations were made in the polar CH on 21 September 1996 from 00:15 to 07:30 UTC. The region studied is shown in Fig. 1A as a large rectangle superimposed on the Fe XII image of EIT (Extreme-Ultraviolet Imaging Telescope) (*31*). To compare the magnetic field with the SUMER images, we used the MDI magnetograms for the field extrapolation over an area of 384″ by 231″, which is a little bigger than the small rectangle (of size 311″ by 201‴) in Fig. 1A. For details on data analysis, see (*32*).

Figure 1B shows a map of B_z, which is obtained by correcting the MDI data for the LOS effect, that is, by dividing through by $\cos 55° = 0.57$, appropriate for the average latitude. Figure 1, C to E, shows the Si II and C IV radiances and Ne VIII Doppler shift. Hassler *et al.* (*23*) found a relation between the Ne VIII Doppler shift and the chromospheric network, as inferred by visual inspection of the Si II radiance patterns, in the sense that the strongest outflow was observed near network lanes and their intersections. How-

[1]Department of Geophysics, Peking University, Beijing 100871, China. [2]Department of Atmospheric Sciences, Peking University, Beijing 100871, China. [3]Max-Planck-Institut für Sonnensystemforschung, 37191 Katlenburg-Lindau, Germany. [4]School of Earth and Space Sciences, University of Science and Technology of China, Hefei, Anhui 230026, China. [5]National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China.

*To whom correspondence should be addressed. E-mail: cytu@public3.bta.net.cn

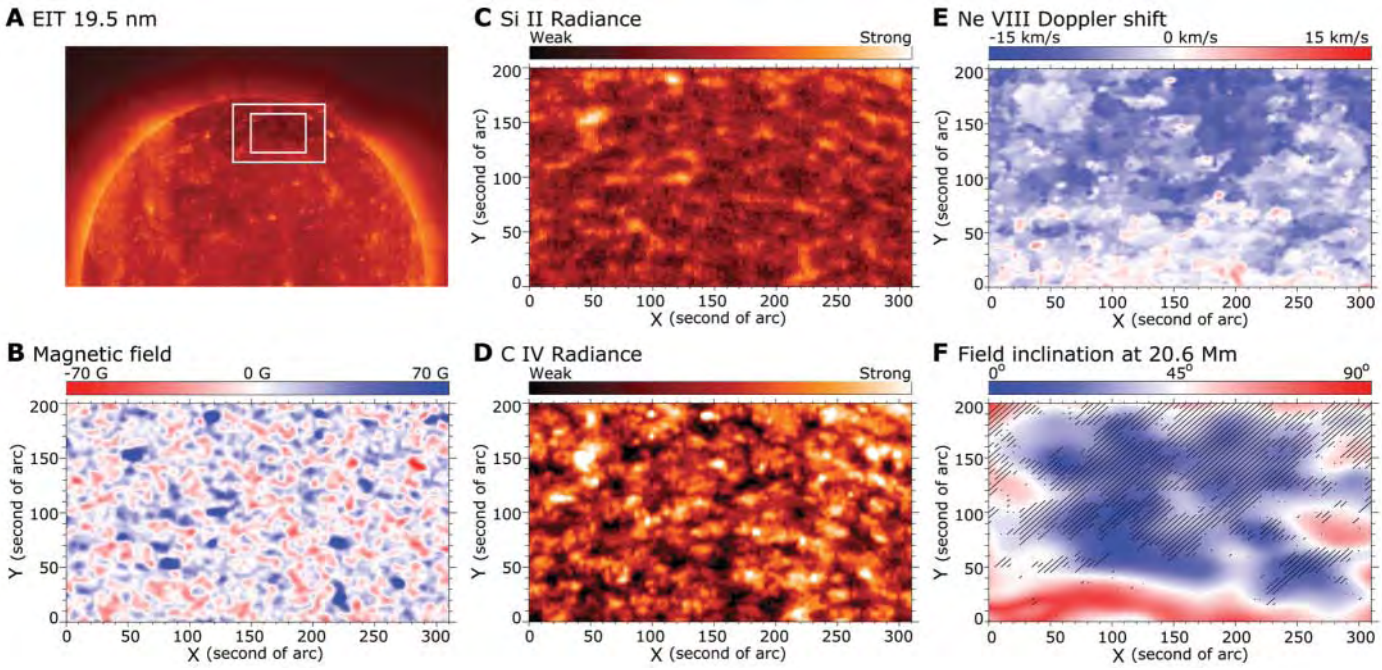


Fig. 1. (**A**) The Sun in the wavelength window around 19.5 nm. The large white rectangle indicates the original size of the SUMER raster scan. A comparison of the structures was made only for the smaller rectangle. (**B**) Magnetic field vertical component ranging from –70 G to 70 G. (**C**) Si II radiance and (**D**) C IV radiance, both in arbitrary units. (**E**) Ne VIII Doppler shifts along the LOS, ranging from –15 km/s to 15 km/s. (**F**) Comparison between the Ne VIII Doppler shift (hatched regions with outflow speeds higher than 7 km/s) and the magnetic field angle, with 0° indicating vertical and 90° horizontal orientations at the height z = 20.6 Mm.

ever, according to Fig. 1, C and E, there is no clear overall correlation, mainly because the structures in the Ne VIII Doppler map are larger than those in the Si II radiance map.

Clearly, the Ne VIII and Si II lines are not emitted at the same height. To determine these heights, we used the method of Seehafer (*33*), as Marsch *et al*. did for ARs [(*29*), equations 1 to 3 for B_x, B_y, and B_z], and extrapolated in a force-free way the photospheric magnetic field to the TR and corona. In the extrapolation box, the positive x coordinate is westward, the y coordinate northward in the tangent plane, and z upward. We considered foreshortening by dividing the apparent distance in the y direction through cos55°. Then we calculated the correlations between radiance maps and 2-D charts of the coronal B_z at different heights.

The result is shown in Fig. 2A. The number of data points for each plot is 62,511. The critical value of 0.008 for the linear correlation coefficient is determined with 95% confidence. The correlation coefficient between the Ne VIII blueshift and $|B_z/B|$ clearly increases with growing z and reaches a maximum of 0.39 at 20.6 Mm. For Si II, it reaches a maximum value of 0.50 at 4.0 Mm and for C IV, 0.18 at 4.5 Mm. The horizontal bar shows the z range in which the coefficient is above 95% of its maximum. The heights of maximal correlation for Si, C, and Ne range from 1.5 to 7.0, 2.0 to 7.0, and 16 to 29 Mm, respectively. For disk observations, the correlation technique is perhaps

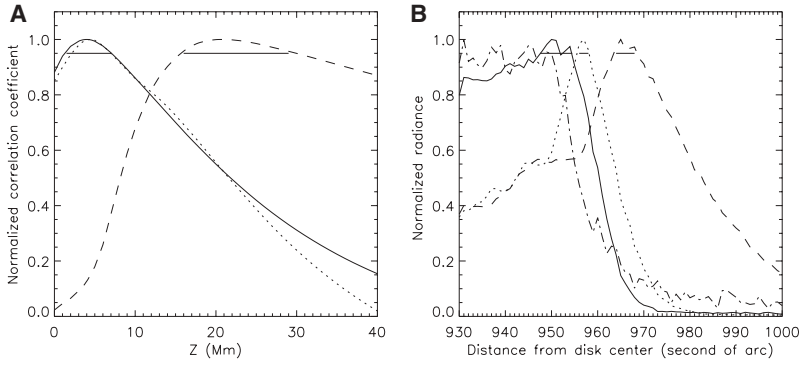


Fig. 2. (**A**) Height variations of the correlation coefficients. The solid line gives the coefficient (normalized by 0.50) between line radiance and $|B_z|$ for Si II, the dotted line the coefficient (normalized by 0.18) for C IV. The dashed line gives the coefficient (normalized by 0.39) between the Ne VIII blueshift and $|B_z/B|$. (**B**) Radial variations of the line radiances of Si II, C IV and Ne VIII, and of the continuum at 154.31 nm, which is indicated by the dash-dot line. The solid line indicates Si II, the dotted C IV, and the dashed Ne VIII. All radiances are normalized to their maximum values. The data were taken across the limb, within the upper right corner of the larger rectangle in Fig. 1A. To indicate the precision by which one can determine the maximum height, we have drawn horizontal bars extending between the two locations where the coefficients drop below 95% of the maximum.

the only quantitative method to determine emission heights. We are convinced that our results are reliable for the following reasons:

(i) Our disk results are consistent with off-limb observations. Figure 2B shows the radial variations of the Si II, C IV, and Ne VIII radiances, with data from the upper right corner of the large rectangle in Fig. 1A. The difference between the maxima of the Si II and Ne VIII profiles is 15″ or 11 Mm (a view angle of 1″ corresponds to a distance of 721 km at the Sun), whereas between the correlation heights it is 16 Mm. Considering the uncertainties of both methods, the results can be considered as consistent.

(ii) Figure 3A shows the Ne VIII Doppler shift versus $|B_z/B|$ at 20.6 Mm. The 31,321 data points were selected under the condition $|B_y/B_x| < 1$ to ensure that the vertical field has a component along the LOS. For a nearly horizontal field, the flow speed is about zero. In contrast, outflows are concentrated at locations of nearly vertical fields. This can be seen in Fig. 1F, which shows a comparison between Ne VIII Doppler shift and field inclination. Figure 3B displays the square

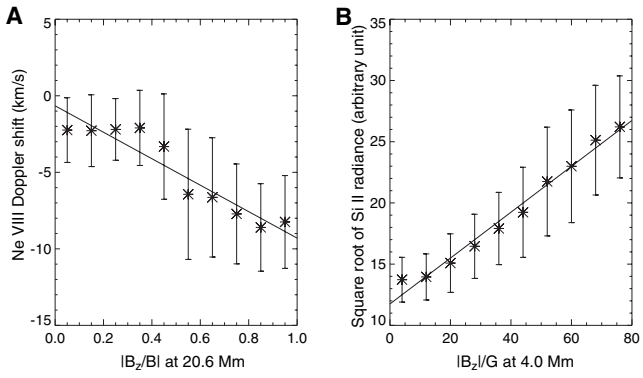


Fig. 3. Results from the linear correlation analyses. The asterisks indicate the average values in vertical bins, with one SD. The fit parameters are given in Table 1. (**A**) Plot of the Ne VIII Doppler shift versus $|B_z/B|$ at 20.6 Mm. (**B**) Plot of the square root of the Si II radiance versus $|B_z|$ at 4.0 Mm.

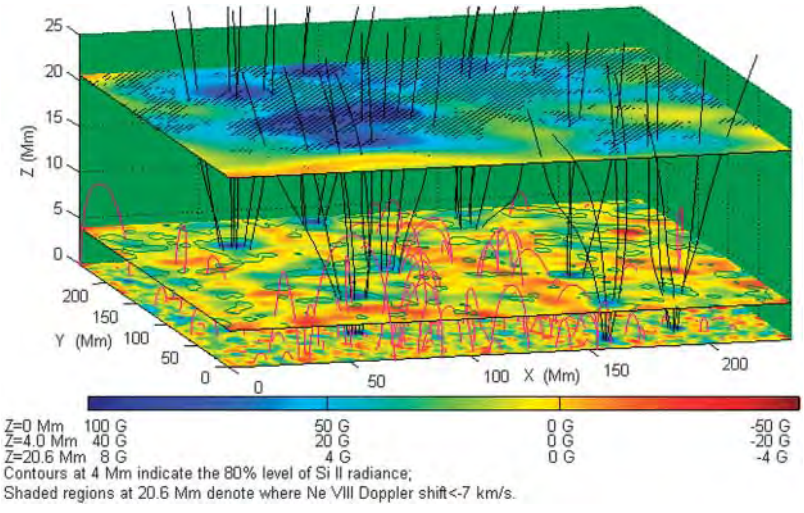


Fig. 4. Magnetic field structures in the 3-D solar atmosphere. The black solid curves illustrate open and the red curves closed field lines. Because the magnetic field strength decreases with increasing height (Z) in the corona, the scales on the color bars differ for different Z. In the plane inserted at 4 Mm, we compare the Si II radiance with the extrapolated B_z. The contours delineate the 80% level of the Si II radiance. In the plane inserted at 20.6 Mm, we compare the Ne VIII Doppler shifts smaller than –7 km/s with the extrapolated B_z. The shaded areas indicate where the Ne^{7+} outflow speed is larger than 7 km/s.

root of the Si II radiance, which is proportional to the electron density, versus the absolute value of B_z at 4.0 Mm. The 39,117 data points were selected under the constraint $|B_z/B| > 0.5$ to ensure that the extrapolated field is vertical. The correlations shown in Fig. 3, A and B, are high. This indicates that the extrapolated field is reasonable and its morphology consistent with the radiance pattern.

(iii) The correlation height hardly changes with α, which has the dimension of an inverse length, the gradient scale of the field. We considered two situations: $\alpha = -1.1 \times 10^{-8}$ m^{-1}, which is the value derived for the force-free field in an AR (*29*); and $\alpha = 0$ for a potential field. Theoretically, steady open flux tubes should have no twist, that is, $\alpha = 0$, because twists will become Alfvén waves escaping from the funnel. We obtained similar correlation heights for both values of α. We confirmed that the field extrapolation in CHs does not critically depend on α up to 40 Mm. The mean heights (variance given in parentheses) of maximum correlation (for $\alpha = 0$) are Si II $z = 4.0$ Mm (1.3–7.1), C IV $z = 5.0$ Mm (2.7–7.5), and Ne VIII $z = 24$ Mm (17.5–33).

(iv) The conditions prevailing in the CH regions fulfill the requirements for a stationary, force-free field extrapolation. Structures in the solar atmosphere such as supergranulation, chromospheric network, and magnetic funnels all are comparatively long-lived (tens of hours), and thus their time variations do not matter. The SUMER data are taken within several hours.

The force-free assumption appears reasonable, considering that the vertical gradient of the thermal pressure balances the gravity force and that the magnetic pressure dominates the thermal pressure. Because the unipolar funnel is surrounded by strong bipolar fields reaching up to about 7 Mm (Fig. 4, lower level), the adjacent loops keep the funnel's cross section constricted, and thus through magnetic tension prevent a rapid horizontal expansion of the funnel with height. This is different from the classical Gabriel model (*19*), in which a horizontal current sheet exists at the bottom of the funnel at 1500 km, where the thermal pressure balances the magnetic pressure. In our model without any current sheets, a unipolar funnel is surrounded by bipolar loops controlling the funnel shape.

The linear force-free field extrapolation may be a good approximation for the TR if the underlying photospheric area is large enough and if no data near the boundaries are used. Our area is 277 Mm by 292 Mm. We considered extrapolation up to 40 Mm and did not use data near boundaries (~27 Mm on each side in the x direction and 11 Mm on each side in the y direction). Short-term magnetic activities such as micro- and nanoflares, or field topology changes and magnetic reconnection can of course not be described by the force-free model.

Outflow pattern and funnel shape. As illustrated in Fig. 3, the Ne^{7+} outflow pattern is closely correlated with the magnetic field inclination $|B_z/B|$. Figure 4 shows a strong correlation with B_z itself. A large B_z and vertical fields represent open coronal funnels, being rooted in strong photospheric fields. Figure 4 shows the field strength at $z = 0$ Mm, $z = 4$ Mm, and $z = 20.6$ Mm, together with Si II radiance contours (top 80%) at $z = 4$ Mm and Ne VIII Doppler shifts at $z = 20.6$ Mm. The field line spreading and associated blue patches indicate the expansion of the funnels. Obviously, the Ne^{7+} outflow is controlled and guided by the funnel shape at 20.6 Mm. The funnels are rooted near supergranulation boundaries in the network lanes.

Where in the funnel does the solar wind acceleration actually start? To answer this question in comparison with models, we identified a single funnel linking the TR with the corona and correlated the observed radiances and Doppler shifts with the extrapolated magnetic fields. Gabriel (*19*) presented his seminal funnel model for a unipolar field region. Dowdy *et al*. (*20*) discussed a static conceptual model for the 3-D geometry of the TR, including multiple loops between open funnels. Other authors (*15*, *17*, *18*) developed various dynamical funnel models based on 1-D fluid equations, including waves and plasma flow.

The extrapolated field in Fig. 4 shows many closed loops between the funnels in this CH region. Some of the loops reach as high as 7 Mm. The funnel rooted at $x = 110$ Mm and $y = 150$ Mm merges with two others, thus forming a large strong-field region. It coincides with the hatched patches, which indicates a comparatively high outflow speed there. Further details of the field geometry around this funnel are shown in Fig. 5A. We only illustrate some open field lines and surrounding closed loops, which seem to quench the funnel.

Figure 5B again shows this funnel, which does not have a canopy shape as originally suggested (*19*). Its cross section increases almost linearly with height, leading to 12 times its photospheric area at 21 Mm. Concurrently, the field strength decreases from 63.9 G

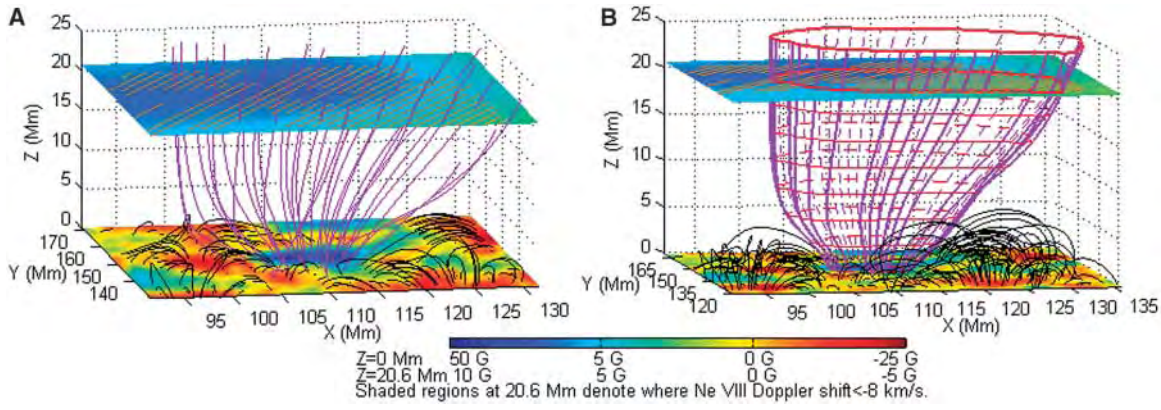


Fig. 5. Magnetic funnel in the solar atmosphere. (**A**) Emphasis on open field lines and correlation with the Ne^{7+} outflow speed larger than 8 km/s (dark shading). (**B**) Illustration of the funnel boundary and magnetic unipolar flux constriction by adjacent, surrounding bipolar loops.

Table 1. Parameters of the linear correlation analysis with 10 magnetic field data bins.

Parameter	Coefficient	$\lvert B_z\rvert$/10 G	$\lvert B_z/B\rvert$
Si II Radiance (arbitrary unit)	*R*	0.99	0.96
	a	11.75 ± 0.41	13.29 ± 0.09
	b	1.87 ± 0.09	1.53 ± 0.16
C IV Doppler shift (km/s)	*R*	0.91	0.92
	a	–0.97 ± 0.11	–0.99 ± 0.08
	b	+0.60 ± 0.10	+0.95 ± 0.15
Ne VIII Doppler shift (km/s)	*R*	–0.76	–0.94
	a	–3.33 ± 1.09	–0.66 ± 0.64
	b	–8.87 ± 2.67	–8.64 ± 1.11

(6.39 mT) in the photosphere to 5.3 G at 21 Mm. How fast a funnel can diverge is determined by the surrounding loops, which push laterally on the funnel. This result agrees with the sketch of intermingling funnels and loops as drawn by Dowdy *et al*. (*20*) and suggested by Peter (*34*, *35*) and supports the idea that "the funnels are constricted at their base in the network not only by the supergranular flow but also by crowding from the ambient magnetic field of the many closed loops" (*20*).

Origin of the solar wind in coronal funnels. To determine where the solar wind acceleration starts in funnels, we statistically evaluated the Doppler shift for regions with strong fields and vertical field lines. In the funnel shown in Fig. 5B, Si^{+} has an average downflow speed of 2.0 ± 1.3 km/s at 4 Mm, C^{3+} an almost zero speed of 0.3 ± 3.3 km/s, and Ne^{7+} a clear upflow speed of 9.6 ± 2.0 km/s. Therefore, no outflow is found for the carbon and silicon ions. For calibration of the Doppler shift, we assumed that its average value is zero just above the solar limb (*36*).

To get firmer results on the flow velocity in the funnels, we calculated average Doppler shifts with data that were above the top 70% of the $|B_z|$ and Si II radiance. This Doppler-shift average indicates downflows of 0.3 km/s (with a σ of 2.8 km/s) for Si^{+} and 0.5 km/s (σ = 3.2 km/s) for C^{3+}. As a result, when using Si^{+} and C^{3+} as flow tracers, no significant solar wind outflow was found.

To compare the Doppler shift in the funnels with that in cell centers, we correlated the Si II, C IV, and Ne VIII Doppler shifts with $|B_z|$ and $|B_z/B|$, at 4.0 Mm, 4.5 Mm, and 20.6 Mm, respectively. It turned out that the downflow of Si^{+} is influenced neither by $|B_z|$ nor by $|B_z/B|$. However, the Doppler shifts of C IV and Ne VIII and the Si II radiance are correlated with these parameters.

The results of the correlation analysis are given in Table 1. The data sets used are averages, derived from 31,321 original data points and obtained by forming 10 bins of $|B_z|$ or $|B_z/B|$. The critical correlation coefficient is 0.58 for a 95% confidence level. In Table 1, *R* is the correlation coefficient, and *a* and *b* are the parameters of the linear correlation $y = a + bx$. The condition of $|B_y/B_x| < 1$ was imposed to guarantee that the field is mainly in the *x*-*z* plane, in which case the LOS Doppler shift indicates vertical motions, because the horizontal magnetic field then has only an *x* component perpendicular to the LOS direction. Consequently, we largely exclude unwanted contributions resulting from Doppler shifts that are caused by horizontal plasma motion directed along the magnetic field. Similarly, the 39,117 points in the data set for Fig. 3B were selected with the requirement that $|B_z/B| > 0.5$, to eliminate data with mostly horizontal fields.

In Table 1, the fit coefficient *b* for C IV is positive. This indicates that downflow of C^{3+} ions may occur in regions with high $|B_z|$ and $|B_z/B|$ and, reversing the argument, that upflow can be expected for low values of $|B_z|$ and $|B_z/B|$. Consistent with supergranular motion, a slight trend appears for downflow near cell boundaries in contrast to upflow in cell centers. No significant outflow can be identified for the C^{3+} ions in the funnels, but for the Ne^{7+} ions we clearly find outflow (together with weak radiance) within the funnels, where both $|B_z|$ and $|B_z/B|$ are large.

Three-dimensional source of the solar wind. Stimulated by our observations, we suggest a way to explain the origin of the fast solar wind. The transition region in CHs is full of magnetic loops of different sizes, mostly with heights of less than 5 Mm. Supergranular plasma convection in the photosphere keeps the feet of the loops moving and thus transfers kinetic energy to magnetic energy that is stored in the loops. They may finally move to a funnel region and undergo reconnection with existing open fields. Thereby, plasma previously confined in the loops is released, which may lead to both upflows and downflows. Ultimately, parts of the plasma contained in reconnecting loops are brought into the corona.

In the lower TR below about 5 Mm, we mainly have horizontal exchange of mass and energy between neighboring flux tubes, which is driven by supergranular motion. Above 5 Mm or higher, where reconnection between field lines of funnels and surrounding loops gradually ceases, vertical transport will become more important than horizontal, and the radial acceleration of the solar wind will actually start. Of course, the subsequent requirements on solar wind heating and acceleration are the same as in standard models.

This scenario is supported by the evidence provided in Fig. 3B, establishing a linear relation between the square root of the Si II radiance (as proxy of the plasma density ρ) and $|B_z|$, and by the observation in Fig. 4 that the highest Si II radiance is concentrated in the strong-field funnels. From magnetohydrodynamic theory we know that, for fields frozen in the plasma, the ratio B/ρ should be conserved (*37*), given that the flow velocity does not vary strongly along **B**. This is the case for horizontal plasma flow at 4 Mm, which is forced by magnetoconvection of a mainly vertical field. Thus, strong Si II emission may indicate mass concentrations caused by accumulation of advected magnetic flux.

Our scenario is akin to the magnetic furnace model proposed by Axford and McKenzie (*14–16*) and to ideas invoking reconnection of mesoscale loops (*38*, *39*). We adopt from the furnace model the idea that reconnection plays a major role, as it will release plasma, set free magnetic energy, and produce Alfvén waves. However, our model of the nascent solar wind is intrinsically 3-D, and the magnetic field geometry is derived empirically. The plasma is accelerated in the funnel above a critical height of 5 Mm but originates below from the neighboring loops. The initial heating of the solar wind plasma is achieved in the side loops.

References and Notes

1. L. Biermann, *Zeitschrift für Astrophysik* **29**, 274 (1951).
2. E. N. Parker, *Astrophys. J.* **128**, 664 (1958).
3. M. Neugebauer, C. W. Snyder, *Science* **138**, 1095 (1962).
4. M. Neugebauer, C. W. Snyder, *J. Geophys. Res.* **71**, 4469 (1966).
5. Y. C. Whang, C. C. Chang, *J. Geophys. Res.* **70**, 4175 (1965).
6. R. E. Hartle, P. A. Sturrock, *Astrophys. J.* **151**, 1155 (1968).
7. A. J. Hundhausen, *Coronal Expansion and Solar Wind* (Springer-Verlag, Heidelberg, 1972).
8. E. Leer, T. E. Holzer, *Solar Phys.* **63**, 143 (1979).
9. E. Leer, T. E. Holzer, *J. Geophys. Res.* **85**, 4681 (1980).
10. C.-Y. Tu, *Solar Phys.* **109**, 149 (1987).
11. J. V. Hollweg, *J. Geophys. Res.* **91**, 4111 (1986).
12. V. H. Hansteen, E. Leer, *J. Geophys. Res.* **100**, 21577 (1995).
13. C.-Y. Tu, E. Marsch, *Solar Phys.* **171**, 363 (1997).
14. J. F. McKenzie, G. V. Sukhorukova, W. I. Axford, *Astron. Astrophys.* **330**, 1145 (1998).
15. W. I. Axford, J. F. McKenzie, in *Solar Wind Seven*, E. Marsch and R. Schwenn, Eds., (Pergamon Press, Oxford, 1992), pp. 1–5.
16. W. I. Axford, J. F. McKenzie, in *Cosmic Winds and the Heliosphere*, J. R. Jokipii, C. P. Sonett, M. S. Giampapa, Eds. (Arizona University Press, Tucson, 1997), pp. 31–66.
17. E. Marsch, C.-Y. Tu, *Solar Phys.* **176**, 87 (1997).
18. P. Hackenberg, E. Marsch, G. Mann, *Astron. Astrophys.* **360**, 1139 (2000).
19. A. H. Gabriel, *Philos. Trans. R. Soc.* **A281**, 339 (1976).
20. J. F. Dowdy Jr., A. G. Emslie, R. L. Moore, *Solar Phys.* **112**, 255 (1987).
21. K. Wilhelm *et al.*, *Solar Phys.* **170**, 75 (1997).
22. P. Lemaire *et al.*, *Solar Phys.* **170**, 105 (1997).
23. D. M. Hassler *et al.*, *Science* **283**, 810 (1999).
24. K. Wilhelm, I. E. Dammasch, E. Marsch, D. M. Hassler, *Astron. Astrophys.* **353**, 749 (2000).
25. L. D. Xia, E. Marsch, W. Curdt, *Astron. Astrophys.* **399**, L5 (2003).
26. L. D. Xia, E. Marsch, K. Wilhelm, *Astron. Astrophys.* **424**, 1025 (2004).
27. T. Wiegelmann, T. Neukirch, *Solar Phys.* **208**, 233 (2002).
28. P. H. Scherrer *et al.*, *Solar Phys.* **162**, 129 (1995).
29. E. Marsch, T. Wiegelmann, L. D. Xia, *Astron. Astrophys.* **428**, 629 (2004).
30. Ø. Lie-Svendsen, V. H. Hansteen, E. Leer, *Astrophys. J.* **596**, 621 (2003).
31. J. P. Delaboudinière *et al.*, *Solar Phys.* **162**, 291 (1995).
32. Materials and methods are available as supporting material on *Science* Online.
33. N. Seehafer, *Solar Phys.* **58**, 215 (1978).
34. H. Peter, *Astron. Astrophys.* **374**, 1108 (2001).
35. H. Peter, B. V. Gudiksen, Å. Nordlund, *Astrophys. J.* **617**, L85 (2004).
36. I. E. Dammasch, K. Wilhelm, W. Curdt, D. M. Hassler, *Astron. Astrophys.* **346**, 285 (1999).
37. T. J. M. Boyd and J. J. Sanderson, *Plasma Dynamics* (Thomas Nelson and Sons, London, 1969).
38. W. C. Feldman, B. L. Barraclough, J. L. Phillips, Y.-M. Wang, *Astron. Astrophys.* **316**, 355 (1996).
39. L. A. Fisk, *J. Geophys. Res.* **108**, 1157 (2003).

40. The National Natural Science Foundation of China supported C.-Y.T., C.Z., and L. Z. under projects with the contract nos. 40336053, 40174045 and 40436015; J.-X.Wang, contract no. 10233050; and L.-D.X., contract no. 40436015. The foundation Major Project of National Basic Research supported C.-Y.T., C.Z., L.-D.X., L. Z., and J.-X.W. under contract no. G-200078405. C.-Y.T. is supported by the Beijing Education Project XK100010404. The SUMER project is financially supported by DLR, CNES, NASA, and the ESA PRODEX program (Swiss contribution). SUMER, EIT, and MDI are instruments on SOHO, an ESA and NASA mission. We thank the teams of MDI and EIT for providing the magnetic field data and the context image.

Supporting Online Material

www.sciencemag.org/cgi/content/full/308/5721/519/DC1
Materials and Methods
Fig. S1
References

6 January 2005; accepted 1 March 2005
10.1126/science.1109447

Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data

Karen Sachs,[1]* Omar Perez,[2]* Dana Pe'er,[3]* Douglas A. Lauffenburger,[1]† Garry P. Nolan[2]†

Machine learning was applied for the automated derivation of causal influences in cellular signaling networks. This derivation relied on the simultaneous measurement of multiple phosphorylated protein and phospholipid components in thousands of individual primary human immune system cells. Perturbing these cells with molecular interventions drove the ordering of connections between pathway components, wherein Bayesian network computational methods automatically elucidated most of the traditionally reported signaling relationships and predicted novel interpathway network causalities, which we verified experimentally. Reconstruction of network models from physiologically relevant primary single cells might be applied to understanding native-state tissue signaling biology, complex drug actions, and dysfunctional signaling in diseased cells.

Extracellular cues trigger a cascade of information flow, in which signaling molecules become chemically, physically, or locationally modified; gain new functional capabilities; and affect subsequent molecules in the cascade, culminating in a phenotypic cellular response. Mapping of signaling pathways typically has involved intuitive inferences arising from the aggregation of studies of individual pathway components from diverse experimental systems. Although pathways are often conceptualized as distinct entities responding to specific triggers, it is now understood that interpathway cross-talk and other properties of networks reflect underlying complexities that cannot be explained by the consideration of individual pathways or model systems in isolation. To properly understand normal cellular responses and their potential disregulation in disease, a global multivariate approach is required (*1*). Bayesian networks (*2*), a form of graphical models, have been proffered as a promising framework for modeling complex systems such as cell signaling cascades, because they can represent probabilistic dependence relationships among multiple interacting components (*3–5*). Bayesian network models illustrate the effects of pathway components on each other (that is, the dependence of each biomolecule in the pathway on other biomolecules) in the form of an influence diagram. These models can be automatically derived from experimental data through a statistically founded computational procedure termed network inference. Although the relationships are statistical in nature, they can sometimes be interpreted as causal influence connections when interventional data are used; for example, with the use of kinase-specific inhibitors (*6*, *7*).

There are several attractive properties of Bayesian networks for the inference of signaling pathways from biological data sets. Bayesian networks can represent complex stochastic nonlinear relationships among multiple interacting molecules, and their probabilistic nature can accommodate noise that is inherent to biologically derived data. They can describe direct molecular interactions as well as indirect influences that proceed through additional unobserved components, a property crucial for discovering previously unknown effects and unknown components. Therefore, very complex relationships that likely exist in

[1]Biological Engineering Division, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA. [2]Stanford University School of Medicine, The Baxter Laboratory of Genetic Pharmacology, Department of Microbiology and Immunology, Stanford, CA 94305, USA. [3]Harvard Medical School, Department of Genetics, Boston, MA 02115, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: lauffen@mit.edu (D.A.L.); gnolan@stanford.edu (G.P.N.)

signaling pathway architectures can be modeled and discovered. The Bayesian network inference algorithm constructs a graph diagram in which nodes represent the measured molecules, and arcs (drawn as lines between nodes) represent statistically meaningful relations and dependencies between these molecules. When inferring a Bayesian network from experimental data, the network inference algorithm aims to discern a model that closely predicts the observations made. The algorithm approximates the most likely models by traversing the space of possibilities via single-arc changes that improve the score. There is a trade-off between simple models and those that accurately capture the empirical distribution observed in the data. The employed Bayesian scoring metric captures this trade-off; thus, a high-scoring model is a both simple and accurate representation of the data (*8*, *9*). Bayesian networks have been applied to gene expression data for the study and discovery of genetic regulatory pathways (*4*, *6*, *10*). However, because of the probabilistic nature of the Bayesian modeling approach, effective inference requires many observations of the system. Thus, such studies have often been limited by data sets of insufficient size; for instance, those made up of measurements based on averaged samples derived from heterogeneous cell populations (a necessary limitation when using lysates from large numbers of cells) (*5*, *11*).

In contrast to lysate-based methods, intracellular multicolor flow cytometry (*12*, *13*) allows more quantitative simultaneous observations of multiple signaling molecules in many thousands of individual cells. Hence, it is an especially appropriate source of data for Bayesian network modeling of signaling pathways; for instance, because it allows for simultaneous measurement of biological states in more native contexts, as well as for large sample sets. Flow cytometry can be used to quantitatively measure a given protein's expression level and can also include measures of protein-modification states such as phosphorylation (*13*–*15*). Because each cell is treated as an independent observation, flow cytometric data provide a statistically large sample that could enable Bayesian network inference to accurately predict pathway structure (Fig. 1A). As demonstrated in this article, interrogating signaling networks in populations of single cells provides a robust source of statistically powerful dependencies that can be used to automatically infer signaling causality using Bayesian network computation.

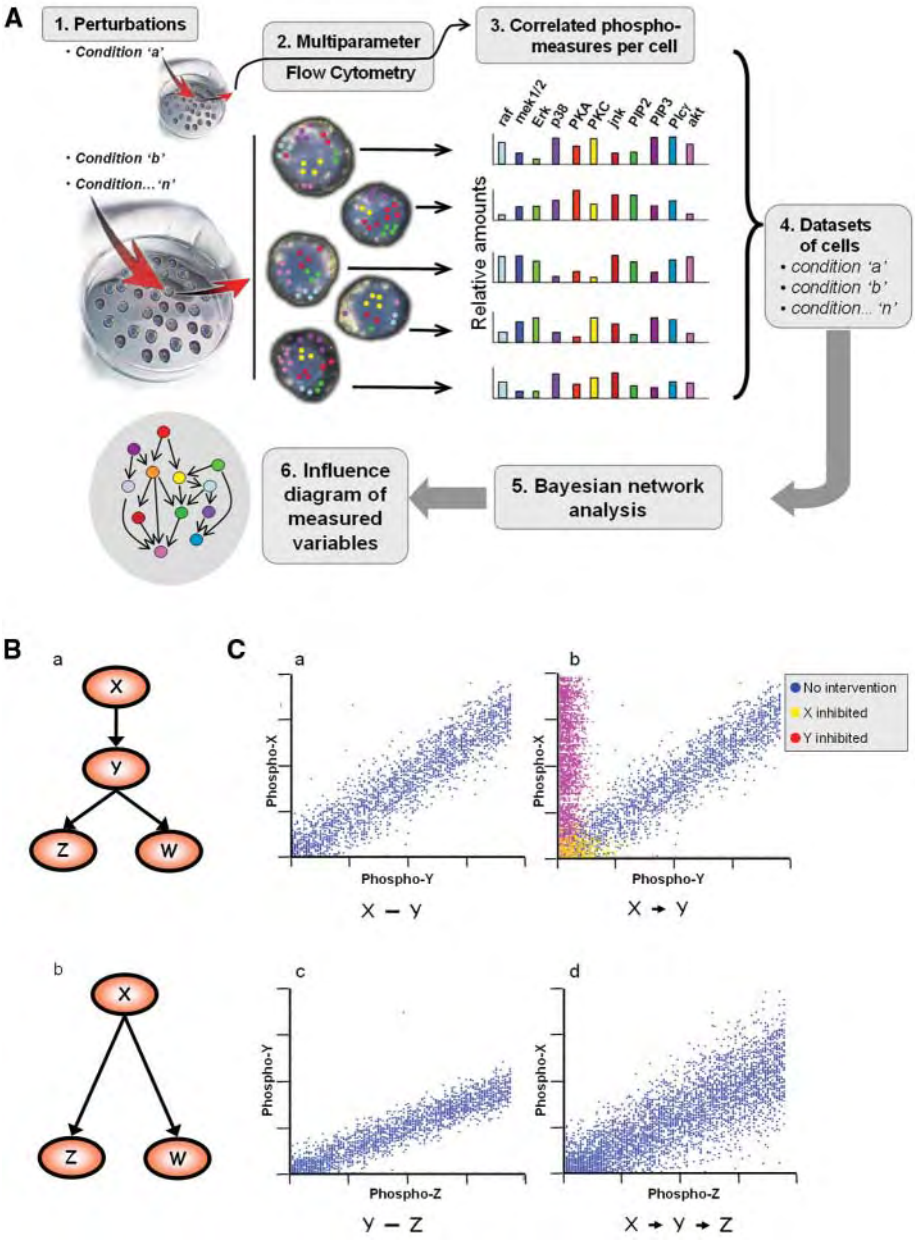


Fig. 1. Bayesian network modeling with single-cell data. (**A**) Schematic of Bayesian network inference using multidimensional flow cytometry data. Nine different perturbation conditions were applied to sets of individual cells (Table 1). A multiparameter flow cytometer simultaneously recorded levels of 11 phosphoproteins and phospholipids in individual cells in each perturbation data set (Table 2). This data conglomerate was subjected to Bayesian network analysis, which extracts an influence diagram reflecting dependencies and causal relationships in the underlying signaling network. (**B**) Bayesian networks for hypothetical proteins X, Y, Z, and W. (a) In this model, X influences Y, which, in turn, influences both Z and W. (b) The same network as (a), except that Y was not measured in the data set. (**C**) Simulated data that could reconstruct the influence connections in (B) (this is a simplified demonstration of how Bayesian networks operate). Each dot in the scatter plots represents the amount of two phosphorylated proteins in an individual cell. (a) Scatter plot of simulated measurements of phosphorylated X and Y shows correlation. (b) Interventional data determine directionality of influence. X and Y are correlated under no manipulation (blue dots). Inhibition of X affects Y (yellow dots) and inhibition of Y does not affect X (red dots). Together, this indicates that X is consistent with being an upstream parent node. (c) Simulated measurements of Y and Z. (d) A noisy but distinct correlation is observed between simulated measurements of X and Z.

Modeling Bayesian networks with multivariable individual-cell data. Fig. 1B (panel a) presents a sample Bayesian network representing four hypothetical biomolecules. A directed arc from X to Y is interpreted as a causal influence from X onto Y; in this case, we say X is Y's "parent" in the network. In the case that X activates Y, where activation can be read out by phosphorylation status, we expect and observe correlation in levels of phosphorylation as measured by flow cytometry (simulated data in Fig. 1C, panel a). Critical to causal interpretation of Bayesian network models is the inclusion of interventional cues (whether activating or inhibiting) that directly

perturb the states of the measured molecules (Fig. 1C, panel b) and strengthen inference directionality. For instance, inhibition of molecule X might lead to inhibition of both X and Y, whereas inhibition of molecule Y leads only to inhibition of Y. Thus, we would infer X to be upstream of Y as shown in Fig. 1B, panel a. Moreover, because flow cytometry can measure multiple molecules within each cell, it is possible to identify complex causal influence relationships involving multiple proteins. Consider the signaling cascade from X onto Y onto Z (Fig. 1B, panel a), where correlation exists between the measured activities of each pair, including between X and Z (Fig. 1C, panel d). Bayesian network inference yields the most concise model, automatically excluding arcs based on dependencies already explained by the model. Thus, despite the correlation between them, the arc between X and Z is omitted, because the X-Y and the Y-Z relationships explain the X-Z correlation. Similarly, because Z and W are both activated by their common cause Y, we expect their activities to be correlated, but no arc appears between them because their respective arcs from Y mediate this dependency. Finally, consider a scenario in which molecule Y was not measured. The statistical correlation between the observed activities of X and Z does not depend on observing Y; therefore, their correlation would still be detected. An indirect arc would be detected from X onto Z (Fig. 1B, panel b).

Expanding this concept to a real data set, we applied Bayesian network analysis to multivariate flow cytometry data. Data were collected after a series of stimulatory cues and inhibitory interventions (Table 1), with cell reactions stopped at 15 min after stimulation by fixation, to profile the effects of each condition on the intracellular signaling networks of human primary naïve CD4+ T cells, downstream of CD3, CD28, and LFA-1 activation (Fig. 2 shows a currently accepted consensus network). We made flow cytometry measurements of 11 phosphorylated proteins and phospholipids [Raf phosphorylated at position S259, mitogen-activated protein kinases (MAPKs) Erk1 and Erk2 phosphorylated at T202 and Y204, p38 MAPK phosphorylated at T180 and Y182, Jnk phosphorylated at T183 and Y185, AKT phosphorylated at S473, Mek1 and Mek2 phosphorylated at S217 and S221 (both isoforms of the protein are recognized by the same antibody), phosphorylation of protein kinase A (PKA) substrates [cAMP response element–binding protein (CREB), PKA, calcium/calmodulin-dependent protein kinase II (CaMKII), caspase-10, and caspase-2] containing a consensus phosphorylation motif, phosphorylation of phospholipase C–γ (PLC-γ) on Y783, phosphorylation of PKC on S660, phosphatidylinositol 4,5-bisphosphate (PIP_2), and phosphatidylinositol 3,4,5-triphosphate (PIP_3)] (Table 2) (*8*, *16*). Each independent

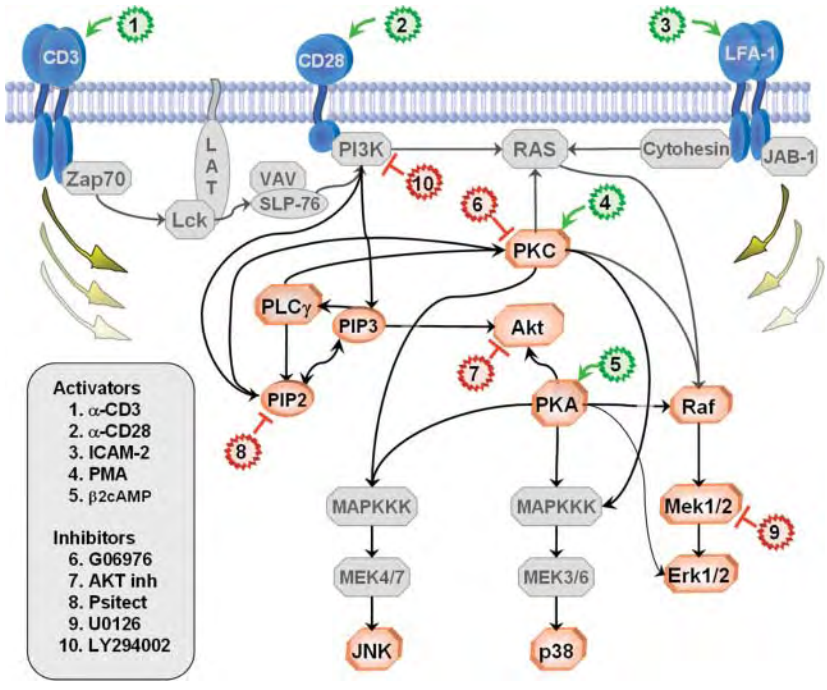


Fig. 2. Classic signaling network and points of intervention. This is a graphical illustration of the conventionally accepted signaling molecule interactions, the events measured, and the points of intervention by small-molecule inhibitors. Signaling nodes in color were measured directly. Signaling nodes in gray were not measured, but are presented to place the signaling nodes that were measured within contextual cellular pathways. The interventions classified as activators are colored green and inhibitors are colored red. Intervention site of action is indicated in the figure. Arcs are used to illustrate connections between signaling molecules; in some cases, the connections may be indirect and may involve specific phosphorylation sites of the signaling molecules (see Table 3 for details of these connections). This figure contains a synopsis of signaling in mammalian cells and is not representative of all cell types, with inositol signaling corelationships being particularly complex.

Table 1. Known biological effects of perturbations employed. The left-hand column lists the specific reagents used in each perturbation condition, and the right-hand column classifies the reagent class into either a general perturbation that overall stimulated the cell or a specific perturbation that acted on a defined set of molecules. The conditions used in the study were as follows: (i) anti-CD3 + anti-CD28, (ii) anti-CD3/CD28 + ICAM-2 (intercellular adhesion molecule–2), (iii) anti-CD3/CD28 + U0126, (iv) anti-CD3/CD28 + AKT inhibitor, (v) anti-CD3/CD28 + G06976, (vi) anti-CD3/CD28 + psitectorigenin, (vii) anti-CD3/CD28 + LY294002, (viii) phorbol 12-myristate 13-acetate (PMA), and (ix) β2 cyclic adenosine 3′,5′-monophosphate (β2cAMP).

Reagent	Reagent class
Anti-CD3/CD28	General perturbation: Activates T cells and induces proliferation and cytokine production. Induced signaling through the T cell receptor (TCR), activated ZAP70, Lck, PLC-γ, Raf, Mek, Erk, and PKC. The TCR signaling converges on transcription factors NFκB, NFAT, and AP-1 to initiate IL-2 transcription.
ICAM-2	General perturbation: Induces LFA-1 signaling and contributes to CD3/CD28 signaling that converges on AP-1 and NFAT transcriptional activity.
β2cAMP	Specific perturbation: cAMP analog that activates PKA. PKA can regulate NFAT activation and T cell commitment processes.
AKT inhibitor	Specific perturbation: Binds inositol pleckstrin domain of AKT and blocks AKT translocation to the membrane where normally AKT becomes phosphorylated and active [median inhibitory concentration (IC_{50}) = 5 mM]. Inhibition of AKT and phosphorylation of AKT substrates are needed to enhance cell survival.
U0126	Specific perturbation: Inhibits MEK1 (IC_{50} = 72 nm) and MEK2 (IC_{50} = 58 nm) in a noncompetitive manner (ATP and Erk substrates). Inhibits activation of Erk, arresting T cell proliferation and cytokine synthesis.
PMA	Specific perturbation: PMA activates PKC and initiates some aspects of T cell activation.
G06976	Specific perturbation: Inhibits PKC isozymes (IC_{50} < 8 nM). Inhibits PKC and arrests T cell activation.
Psitectorigenin	Specific perturbation: Inhibits phosphoinositide hydrolysis. Inhibits PIP2 production and disrupts phosphoinositol turnover.
LY294002	Specific perturbation: Phosphatidylinositol 3-kinase (PI3K inhibitor. Inhibits PI3K and subsequent activation of AKT.

sample in this data set consists of quantitative amounts of each of the 11 phosphorylated molecules, simultaneously measured from single cells [data sets are downloadable (*8*)]. For purposes of illustration, examples of actual fluorescence-activated cell sorter (FACS) data plotted in prospective corelationship form are shown in fig. S1. In most cases, this reflects the activation state of the kinases monitored, or in the cases of PIP_3 and PIP_2, the levels of these secondary messenger molecules in primary cells, under the condition measured. Nine stimulatory or inhibitory interventional conditions were used (Table 1) (*8*). The complete data sets were analyzed with the Bayesian network structure inference algorithm (*6*, *9*, *17*).

A high-accuracy human primary T cell signaling causality map. The resulting de novo causal network model was inferred (Fig. 3A) with 17 high-confidence causal arcs between various components. To evaluate the validity of this model, we compared the model arcs (and absent potential arcs) with those described in the literature. Arcs were categorized as the following: (i) expected, for connections well-established in the literature that have been demonstrated under numerous conditions in multiple model systems; (ii) reported, for connections that are not well known, but for which we were able to find at least one literature citation; and (iii) missing, which indicates an expected connection that our Bayesian network analysis failed to find. Of the 17 arcs in our model, 15 were expected, all 17 were either expected or reported, and 3 were missed (Fig. 3A and table S1) (*8*, *18*–*22*). Table 3 enumerates the probable paths of influence corresponding to model arcs determined by surveying published reports.

Several of the known connections from our model are direct enzyme-substrate relationships (Fig. 3B) (PKA to Raf, Raf to Mek, Mek to Erk, and Plc-γ to PIP_2), and one has a relationship of recruitment leading to phosphorylation (Plc-γ to PIP_3). In almost all cases, the direction of causal influence was correctly inferred (an exception was Plc-γ to PIP_3, in which case the arc was inferred in the reverse direction). All the influences are contained within one global model; thus, the causal direction of arcs is often compelled so that these are consistent with other components in the model. These global constraints allowed detection of certain causal influences from molecules that were not perturbed in our assay. For instance, although Raf was not perturbed in any of the measured conditions, the method correctly inferred a directed arc from Raf to Mek, which was expected for the well-characterized Raf-Mek-Erk signal transduction pathway. In some cases, the influence of one molecule on another was mediated by intermediate molecules that were not measured in the data set. In the results, these indirect connections were detected as well (Fig. 3B, panel b). For example, the influence of PKA and PKC on the MAPKs p38 and Jnk likely proceeded via their respective (unmeasured) MAPK kinase kinases. Thus, unlike some other approaches used to elucidate signaling networks [for example, protein-protein interaction maps (*23*, *24*)] that provide static biochemical association maps with no causal links, our Bayesian network method can detect both direct and indirect causal connections and therefore provide a more contextual picture of the signaling network.

Another feature demonstrated in our model is the ability to dismiss connections that are already explained by other network arcs (Fig. 3B, panel c). This is seen in the Raf-Mek-Erk cascade. Erk, also known as p44/42, is downstream of Raf and therefore dependent on Raf, yet no arc appears from Raf to Erk, because the connection from Raf to Mek and the connection from Mek to Erk explain the dependence of Erk on Raf. Thus, an indirect arc should appear only when one or more intermediate molecules is not present in the data set, otherwise the connection will proceed via this molecule. The intervening molecule may also be a shared parent. For example, the phosphorylation statuses of p38 and Jnk are correlated (fig. S2), yet they are not directly connected, because their shared parents (PKC and PKA) mediate the dependence between them. Although we cannot know whether an arc in our model represents a direct or indirect influence, it is unlikely that our model contains an indirect arc that is mediated by any molecule observed in our measurements. Correlation exists between most molecule pairs in this data set [per Bonferroni corrected *P* value (fig. S2)], which can occur with closely connected pathways. Therefore, the relative lack of arcs in our model (Fig. 3A) contributed greatly to the accuracy and interpretability of the inferred model.

A more complex example is the influence of PKC on Mek, which is known to be mediated by Raf (Fig. 3B, panel d). PKC is known to affect Mek through two paths of influence, each mediated by a different active phosphorylated form of the protein Raf. Although PKC phosphorylates Raf directly at S499 and S497, this event is not detected by our measurements, because we use only an antibody specific to Raf phosphorylation at S259 (Table 2) (*16*). Therefore, our algorithm detects an indirect arc from PKC to Mek that is mediated by the presumed unmeasured intermediate Raf phosphorylated at S497 and S499 (*18*). The PKC-to-Raf arc represents an indirect influence that proceeds via an unmeasured molecule, presumed to be Ras (*19*, *20*). We discussed above the ability of our approach to dismiss redundant arcs. In this case, there are two paths leading from PKC to Mek, because each path corresponds to a separate means of influence from PKC to Mek: one via Raf phosphorylated at S259 and the other through Raf phosphorylated at S497 and S499. Thus, neither path is redundant. This result demonstrates the distinction that this analysis is sensitive to specific phos-

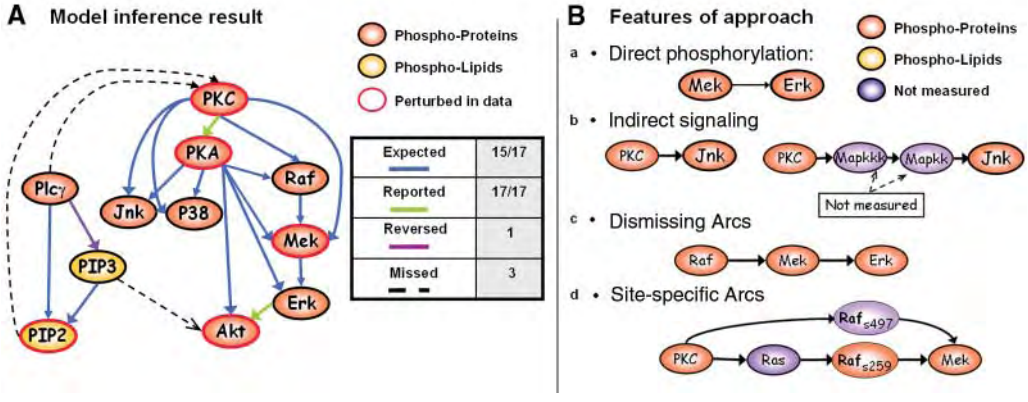


Fig. 3. Bayesian network inference results. (**A**) Network inferred from flow cytometry data represents expected outcomes. This network represents a model average from 500 high-scoring results. High-confidence arcs, appearing in at least 85% of the networks, are shown. For clarity, the names of the molecules are used to represent the measured phosphorylation sites (Table 2). (**B**) Inferred network demonstrates several features of Bayesian networks. (a) Arcs in the network may correspond to direct events or (b) indirect influences. (c) When intermediate molecules are measured in the data set, indirect influences rarely appear as an additional arc. No additional arc is added between Raf and Erk because the dependence between Raf and Erk is dismissed by the connection between Raf and Mek, and between Mek and Erk (for instance, see Fig. 1C). (d) Connections in the model contain phosphorylation site–specificity information. Because Raf phosphorylation on S497 and S499 was not measured in our data set, the connection between PKC and the measured Raf phosphorylation site (S259) is indirect, likely proceeding via Ras. The connection between PKC and the undetected Raf phosphorylation on S497 and S499 is seen as an arc between PKC and Mek.

phorylation sites on molecules and is capable of detecting more than one route of influence between molecules.

Three well-established influence connections do not appear in our model: PIP_2 to PKC, PLC-γ to PKC, and PIP_3 to Akt. Bayesian networks are constrained to be acyclic, so if the underlying network contains feedback loops, we cannot necessarily expect to uncover all connections (fig. S3). Availability of suitable temporal data could possibly permit this limitation to be overcome using dynamic Bayesian networks (*25*, *26*).

Experimental confirmation of predicted network causality. Two influence connections in our model are not well established in the literature: PKC on PKA and Erk on Akt. To probe the validity of these proposed causal influences, we searched for reports in the literature. Both connections have previously been reported: the PKC-to-PKA connection in rat ventricular myocytes and the Erk-to-Akt connection in colon cancer cell lines (*21*, *22*). An important goal of our work was to test the ability of Bayesian network analysis of flow cytometry data to correctly infer causal influences from unperturbed molecules within a network. For example, Erk was not directly acted on by any activator or inhibitor in the sample sets, yet Erk showed an influence connection to Akt. Our model thus predicts that direct perturbation of Erk would influence Akt (Fig. 4A). On the other hand, although Erk and PKA are correlated (fig. S2), the model predicts that perturbation of Erk should not influence PKA.

As a test of these predictions (Fig. 3A), we used small interfering RNA (siRNA) inhibition of either Erk1 or Erk2, and the amounts of S473-phosphorylated Akt and phosphorylated PKA were then measured. In accord with the model predictions, Akt ($P < 9.4 \times 10^{-5}$) phosphorylation was reduced after siRNA inhibition of Erk1 but the activity of PKA ($P < 0.28$) was not (Fig. 3, B and C). Akt phosphorylation was not affected by the inhibition of Erk2. The connection between Erk1 and Akt may be direct or indirect, involving mediatory molecules yet to be understood, but the connection is supported by both the model and the validation experiment.

Enablers of accurate inference: network interventions and sufficient numbers of single cells. Three features distinguish our data from the majority of currently attainable biological data sets. First, we simultaneously measured multiple protein states in individual cells, eliminating population-averaging effects that could obscure interesting correlations. Second, because the measurements were on single cells, thousands of data points were collected in each experiment. This feature constitutes a tremendous asset for Bayesian network modeling, because the large number of observations allows for accurate assessment of underlying probabilistic relationships, and therefore allows for the extraction of complex relationships from noisy data. Third, interventional assays generated hundreds of individual data points per intervention (because flow cytometry measures single cells in population), allowing for an increase in inferences of causality. To evaluate the importance of these features, we created the following variations on our original data set: (i) an observation-only data set (that is, without any interventional data) of 1200 data points; (ii) a population-averaged (that is, a simulated Western blot) data set; and (iii) a truncated individual-cell data set of size comparable to the simulated Western blot data set (that is, the original data set with most of the data randomly excluded to reduce its size) (*8*).

Bayesian network inference was performed on each set of data. The network inferred from 1200 observational data points included only 10 arcs, all undirected, of which 8 were expected or reported, and 10 arcs were missing (fig. S4A). This result demonstrates that interventions are critical for effective inference, particularly to establish directionality of the connections (Fig. 1B). The truncated single-cell data set (420 data points) shows a large (11-arc) decline in accuracy, missing more connections and reporting more unexplained arcs than its larger (5400 data points) counterpart (fig. S4B). This result emphasizes the importance of sufficiently large data set size in network inference. The network inferred from averaged data (fig. S4C) shows a further five-arc decline in accuracy relative to that inferred from an equal number of single-cell data points, emphasizing the importance of single-cell data. The fact that population averaging destroys some of the signals present in the data may reflect the presence of heterogeneous cellular subsets that are masked by averaging techniques.

Discussion and summary. As shown, we correctly reverse-engineered and rapidly inferred the basic structure of a classically understood signaling network that connects a number of key phosphorylated proteins in human T cell signaling, a map built by classical biochemistry and genetic analysis over the past two decades. The network was automatically constructed with no a priori knowledge of pathway connectivity. The application of Bayesian networks to single-cell flow cytometry has distinct advantages, including an ability to measure events in primary cells after in vivo interventions (thus measuring context-specific signaling biology in tissues), inference of directed arcs and causal-

Table 2. Nodes measured in pathway and specificity antibodies used. The left-hand column shows target molecules measured in this study that were assayed using monoclonal antibody to the target residues (site of phosphorylation or phosphorylated product as described) (*16*).

Measured molecule	Antibody specificity
Raf	Phosphorylation at S259
Erk1 and Erk2	Phosphorylation at T202 and Y204
p38	Phosphorylation at T180 and Y182
Jnk	Phosphorylation at T183 and Y185
AKT	Phosphorylation at S473
Mek1 and Mek2	Phosphorylation at S217 and S221
PKA substrates	Detects proteins and peptides containing a phospho-Ser/Thr residue with arginine at the –3 position
PKC	Detects phosphorylated PKC-α, -βI, -βII, -δ, -ε, -η, and -θ isoforms only at C-terminal residue homologous to S660 of PKC-βII
PLC-γ	Phosphorylation at Y783
PIP_2	Detects PIP_2
PIP_3	Detects PIP_3

Table 3. Possible molecular pathways of influence represented by arcs in the model. Shown are the possible pathways of influence inferred from the data, with the connection shown in Fig. 3A and the unmeasured molecules (in bold) that might mediate indirect influences. E, expected; R, reported. See main text for further discussion. Specific phosphorylation sites are included as subscripts. See table S1 for citations that support the inferences.

Connection	Influence path	Type	Category
PKC→Raf	PKC→**Ras**→Raf_{S259}	Indirect	E
PKC→Mek	PKC→**$Raf_{S497/S499}$**→Mek	Indirect	E
PKC→Jnk	PKC→→**MKKs**→Jnk	Indirect	E
PKC→p38	PKC→→**MKKs**→p38	Indirect	E
PKC→PKA	PKC→**cAMP**→PKA	Indirect	R
PKA→Raf	PKA→Raf_{S259}	Direct	E
PKA→Mek	PKA→**Raf_{S621}**→Mek	Indirect	E
PKA→Erk	PKA→**HePTP**→Erk	Indirect	E
PKA→Jnk	PKA→→**MKKs**→Jnk	Indirect	E
PKA→p38	PKA→→**MKKs**→p38	Indirect	E
Raf→Mek	Direct phosphorylation	Direct	E
PKA→Akt	PKA→**CaMKK**→**Akt_{T308}**→Akt_{S473}	Indirect	E
Mek→Erk	Direct phosphorylation	Direct	E
Plc-γ→PIP_2	Direct hydrolysis to IP3	Direct	E
Plc-γ→PIP_3	Recruitment leading to phosphorylation	Reversed	E
PIP_3→PIP_2	Precursor-product		E
Erk→Akt	Direct or indirect		R

ity therein, and the ability to detect indirect as well as direct connections. This latter point is a powerful feature when the known list of participating molecules may not be exhaustive, and can be especially important when networks are used to assess the effects of system perturbations (as in a pharmaceutical context). A limiting step in the experiment is the availability of suitable reagents; currently, there are about 80 antibodies to phosphorylated molecules that are compatible with flow cytometry, but this number is expected to rapidly increase (*27*, *28*).

Application of this approach to other sets of molecules, cell types, disease states, and interventions (for example, siRNA and dominant negative screens, or pharmaceutical agents) should enhance our understanding of signaling networks, especially with respect to complex nonlinear cross-talk between pathways. Another important experimental issue that this approach can address is the differences among specific primary cell types and cell subpopulations. The traditional understanding of pathway structures as collated from diverse model cell types and organisms demonstrates the essential congruity of basic signaling networks, but does not easily reveal the subtle differences that exist in different primary cell subtypes. It is now possible to appreciate pathway intricacies in primary cell subsets, including those with previously uncharacterized signaling molecules. The application of this approach during biochemical interrogation of cellular subset–specific signaling networks in the course of a disease state or in the presence of pharmaceutical agents can potentially provide important mechanistic information of clinical relevance. For example, this method could identify sets of signaling molecules that explain differences between responses to chemotherapy in patients with cancer (*15*).

Concerning the computational aspect, a key advantage of Bayesian networks is that they are relatively robust to the existence of unobserved variables; for example, their ability to detect indirect influences via unmeasured molecules. At the forefront of Bayesian network research is the development of methods to automatically infer the existence and location of such hidden variables. Although our results were restricted to 11 phosphomolecular measurements per cell, the number of simultaneous parameters measured by flow cytometry is steadily growing (*27*, *28*). As measurement systems improve, and the ability to readily and accurately measure greater numbers of internal signaling events increases, additional opportunities to discover novel influences and pathway structures become possible.

One of the caveats in the use of Bayesian networks for the elucidation of signaling pathways is that they are restricted to be acyclic, whereas signaling pathways are known to be rich in feedback loops. Indeed, our inference missed three classic arcs, most likely for this reason. Given time series data, dynamic Bayesian networks could potentially capture these feedback loops. To measure the amounts of internal phosphorylated proteins, the cells must be fixed. Therefore, continuous, real-time, simultaneous, multiparameter, single-cell time-series data cannot be collected with the current technology. Because Bayesian networks belong to a more general class of probabilistic graphical models, within the formalism of these models it is possible to develop a model that could handle feedback loops, given a series of static time points using the current technology.

Although there is much to be developed both computationally and experimentally, by extending the concepts derived here it is clear that simultaneous multivariate analysis of biological states in multiple discrete entities, such as cells, offers a useful approach for rapidly deriving signaling network hierarchies and structures. Extension of this approach to biological systems involving multiple cell populations, such as solid tissues and organs, or whole-animal studies such as in whole-body fluorescence imaging of phosphorylation states in staged *Caenorhabditis elegans* or *Drosophila* larva, or thin-slice tissue sections from mammalian organs, could allow automated construction of signaling network influences not only within but also across cell boundaries in an increasing number of physiological contexts.

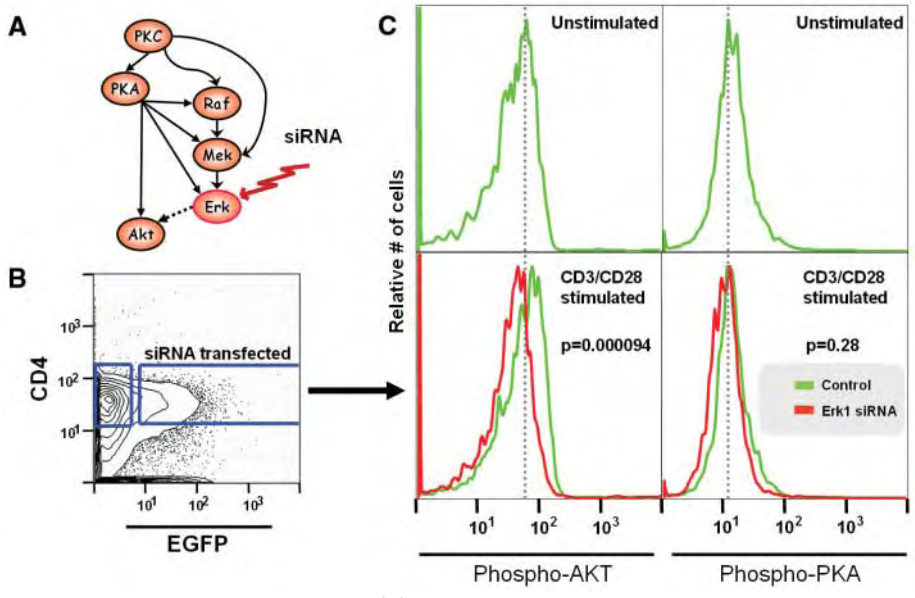


Fig. 4. Validation of model prediction. (**A**) The model predicts that an intervention on Erk will affect Akt, but not PKA. (**B**) To test the predicted relationships, Erk1 and Erk2 were inhibited using siRNA in cells stimulated with antibody to CD3 (anti-CD3) and anti-CD28. (**C**) Amounts of Akt phosphorylation in transfected $CD4^+$ cells [enhanced green fluorescent protein (EGFP+) cells] were assessed, and amounts of phosphorylated PKA are included as a negative control. When Erk1expression is inhibited, phosphorylated Akt is reduced to amounts similar to those in unstimulated cells, confirming our prediction ($P = 0.000094$). PKA is unaffected ($P = 0.28$).

References and Notes

1. T. Ideker, T. Galitski, L. Hood, *Annu. Rev. Genom. Human Gen.* **2**, 343 (2001).
2. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann, San Mateo, CA, 1988).
3. N. Friedman, *Science* **303**, 799 (2004).
4. N. Friedman, M. Linial, I. Nachman, D. Pe'er, *J. Comput. Biol.* **7**, 601 (2000).
5. K. Sachs, D. Gifford, T. Jaakkola, P. Sorger, D. A. Lauffenburger, *Sci. STKE* **2002**, pe38 (2002).
6. D. Pe'er, A. Regev, G. Elidan, N. Friedman, *Bioinformatics* **17** (suppl. 1), S215 (2001).
7. J. Pearl, *Causality: Models, Reasoning, and Inference* (Cambridge Univ. Press, Cambridge, 2000).
8. Materials, methods, and an outline of Bayesian modeling are available as supporting material on *Science* Online.
9. D. Pe'er, *Sci. STKE* **2005**, pl 4 (2005).
10. A. J. Hartemink, D. K. Gifford, T. S. Jaakkola, R. A. Young, *Pac. Symp. Biocomput.* **2001**, 422 (2001).
11. P. J. Woolf, W. Prudhomme, L. Daheron, G. Q. Daley, D. A. Lauffenburger, *Bioinformatics* **21**, 741 (2005).
12. L. A. Herzenberg, D. Parks, B. Sahaf, O. Perez, M. Roederer, *Clin. Chem.* **48**, 1819 (2002).
13. O. D. Perez, G. P. Nolan, *Nat. Biotechnol.* **20**, 155 (2002).
14. O. D. Perez *et al.*, *Nat. Immunol.* **11**, 1083 (2003).
15. J. M. Irish *et al.*, *Cell* **118**, 217 (2004).
16. Single-letter abbreviations for the amino acid residues are as follows: S, Ser; T, Thr; and Y, Tyr.
17. D. Heckerman, in *Learning in Graphical Models*, M. I. Jordan, Ed. (MIT Press, Cambridge, MA, 1998), pp. 301–354.
18. M. P. Carroll, W. S. May, *J. Biol. Chem.* **269**, 1249 (1994).
19. R. Marais, Y. Light, H. F. Paterson, C. J. Marshall, *EMBO J.* **14**, 3136 (1995).
20. R. Marais *et al.*, *Science* **280**, 109 (1998).
21. W. M. Zhang, T. M. Wong, *Am. J. Physiol.* **274**, C82 (1998).
22. R. Fukuda, B. Kelly, G. L. Semenza, *Cancer Res.* **63**, 2330 (2003).
23. M. Steffen, A. Petti, J. Aach, P. D'haeseleer, G. Church, *BMC Bioinformat.* **3**, 34 (2002).
24. B. P. Kelley *et al.*, *Nucleic Acids Res.* **32**, W83 (2004).
25. K. M. Nir Friedman, S. Russell, *Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence* (Morgan Kaufmann, San Francisco, 1998).
26. J. D. G. Irene, M. Ong, D. Page, *Bioinformatics* **18**, S241 (2002).

27. M. Roederer, J. M. Brenchley, M. R. Betts, S. C. De Rosa, *Clin. Immunol.* **110**, 199 (2004).
28. A. Perfetto, P. Chattopadhyay, M. Roederer, *Nat. Rev. Immunol.* **4**, 648 (2004).
29. The authors thank G. Church, N. Friedman, J. Albeck, P. Jasper, L. Garwin, R. Tibshirani, T. Jaakkola, D. Gifford, and D. Koller for helpful discussions or readings of the manuscript and R. Balderas and BD-PharMingen Biosciences for reagents. K.S. and D.A.L. were supported by the National Intitute of General Medical Sciences Center of Excellence in Complex Biomedical Systems at MIT. O.D.P. was supported as a Bristol-Meyer Squibb Irvington Fellow and a Dana Foundation human immunology award. D.P. was supported by a PhRMA Center of Excellence in Integration of Genomics and Informatics grant, an NIH Center of Excellence in Genomic Studies grant to G. Church, and an NSF Postdoctoral Research Fellowship in Biological Informatics. G.P.N. was supported in this work by NIH grants P01-AI39646 and AI35304, a grant from the Juvenile Diabetes Foundation, and National Heart, Lung, and Blood Institute Proteomics Center contract N01-HV-28183I.

Supporting Online Material
www.sciencemag.org/cgi/content/full/308/5721/523/DC
Materials and Methods
Figs. S1 to S4
Table S1
Data Sets S1 to S14
References

29 September 2004; accepted 19 January 2005
10.1126/science.1105809

Parallel and Serial Neural Mechanisms for Visual Search in Macaque Area V4

Narcisse P. Bichot,[1]* Andrew F. Rossi,[2,3] Robert Desimone[1,4]

To find a target object in a crowded scene, a face in a crowd for example, the visual system might turn the neural representation of each object on and off in a serial fashion, testing each representation against a template of the target item. Alternatively, it might allow the processing of all objects in parallel but bias activity in favor of those neurons that represent critical features of the target, until the target emerges from the background. To test these possibilities, we recorded neurons in area V4 of monkeys freely scanning a complex array to find a target defined by color, shape, or both. Throughout the period of searching, neurons gave enhanced responses and synchronized their activity in the gamma range whenever a preferred stimulus in their receptive field matched a feature of the target, as predicted by parallel models. Neurons also gave enhanced responses to candidate targets that were selected for saccades, or foveation, reflecting a serial component of visual search. Thus, serial and parallel mechanisms of response enhancement and neural synchrony work together to identify objects in a scene.

In a crowded visual scene, we typically focus our attention on behaviorally relevant stimuli. When subjects know the location of a relevant object, the brain mechanisms that guide their spatial attention to the object largely overlap with those for selecting the targets for eye movements (*1*). The outcome of this selection for attention or eye movements is to enhance the responses of visual cortex neurons to the relevant object, at the expense of distracters (*2*–*6*). As a result, object recognition mechanisms in the temporal cortex are typically confronted with only a single relevant stimulus at a time (*7*). However, in most common visual scenes, people rarely know the specific location of the relevant object in advance—instead, they must search for it, based on its distinguishing features, such as color or shape, which is commonly termed visual search. A long-standing issue has been whether object selection in visual search is also mediated by neural mechanisms for spatial attention, which scan the objects in the scene sequentially until the target is identified (serial search), whether or not eye movements are made. If so, then visual attention could be broadly served by a unitary mechanism, linked to the neural systems that control gaze. Alternatively, search may be mediated by nonspatial attentional mechanisms that are sensitive to features such as color and shape and that bias visual processing in favor of neurons that represent the target features throughout the visual field, all at once (parallel search) (*7*). Search could also be mediated by hybrid mechanisms such as guided search (*8*).

Previous studies of visual search (*9*–*11*) and attention to stimuli with particular features (*12*–*14*) in brain area V4 have found that neuronal responses to attended target stimuli were enhanced over time, but the studies were not designed to test whether the targets were "found" by serial or parallel neural mechanisms. In one of these studies, monkeys did not search for a specific feature but instead searched for a singleton (i.e., popout) stimulus in one of two feature dimensions (*11*). In another study that used backgrounds of natural scenes, the average neural activity throughout the trial varied according to the searched-for target features, but the authors could not rule out that these effects were due to differences in eye scan paths across the scene for different targets rather than feature-selective effects on neuronal responses (*10*).

We tested for parallel and serial attentional mechanisms in area V4 in monkeys performing a search task with free gaze. We recorded not only neuronal responses but also the synchrony between neuronal responses and the local field potential (LFP) (*15*, *16*), because V4 neurons synchronize their activity when attention is directed to their receptive fields (RFs) (*17*), similar to neurons in parietal cortex during a memory-saccade task (*18*). Such synchrony, especially in the gamma frequency range, could potentially amplify their effect on postsynaptic neurons, similar to increases in firing rate (*19*). The monkeys freely scanned multielement arrays composed of colored shapes to find a target defined by color or shape (*20*). During color feature search (Fig. 1A), the cue was a colored square, and the monkey was rewarded for fixating the stimulus in the array that matched the cue color. During shape feature search (fig. S1), the cue was a gray shape, and the monkeys were rewarded for fixating the stimulus in the array that matched the cue shape. When shape was relevant, color was irrelevant, and vice versa. We selected two colors and two shapes as cues for each recording session, on the basis of initial recordings in which we determined a preferred (strong response) and nonpreferred (weak response) color and shape for a given neuron.

Overall, monkeys performed similarly during color and shape search, finding the target on 86% and 91% of the trials, respectively. Both tasks were demanding, taking an average of 6.3 saccades to find the target out of 20 items (Fig. 1A and fig. S1) (*21*). In separate behavioral studies in which we varied the number of display items, the monkeys took an average of 160 ms per item to find the target, again indicating that the target did not "pop out."

Parallel selection during feature search. The key element of parallel search models is that the neural bias in favor of stimuli containing features of the searched-for target occurs throughout the visual field, and throughout the time period of the search, long before a target is identified. Thus, we reasoned that the critical neurons to test for this bias were not the neurons whose RF contained the stimulus that was the target for a saccade at a given moment. Rather, the critical neurons

[1]Laboratory of Neuropsychology, [2]Laboratory of Brain and Cognition, National Institute of Mental Health (NIMH), National Institutes of Health, Bethesda, MD 20892, USA. [3]Department of Psychology, Vanderbilt University, Nashville, TN 37203, USA. [4]McGovern Institute for Brain Research at the Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

*To whom correspondence should be addressed. E-mail: bichotn@mail.nih.gov

for this test were the ones whose RF contained a potential, or undiscovered, target that was specifically not selected for the next saccade (i.e., not the focus of spatially directed attention) (Fig. 1B). For example, consider two different trials where the animal was cued to search for either a red or a blue target in the color search task. In this case, we would examine the response to a red versus blue stimulus in the RF in the interval when the animal was preparing a saccade to a stimulus somewhere outside the RF (e.g., the orange A in Fig. 1B). The question was whether the response to the red or blue stimulus inside the RF varied according to whether the animal was searching for a red versus blue target, i.e., whether there was any sort of bias in favor of the, as yet undetected, target inside the RF.

For each session and for each neuron, we measured responses under the following four conditions: (i) both the stimulus in the RF and the cue had the neuron's preferred feature (e.g., the animal was searching for red, the neuron preferred red, and there was a red stimulus in the RF); (ii) the RF stimulus had the preferred feature but the cue was the neuron's nonpreferred feature; (iii) the RF stimulus had the nonpreferred feature but the cue was the preferred feature; and (iv) both the RF stimulus and the cue had the nonpreferred feature (*22*).

We used multichannel drives to collect data from 79 single neurons and 70 LFPs in 27 feature search sessions in two monkeys (*23*). The data from the two animals have been combined because they were qualitatively similar. We used population, rather than individual neuron, statistics because of the variable number of stimulus conditions generated by the animal's scanning strategy in a given session (*20*). The average normalized LFP and neuronal responses of the population of V4 neurons for each of the four conditions are shown in Fig. 2 for both color and shape searches (*24*), and the two tasks gave qualitatively similar results. For statistical comparisons, we averaged spike densities and LFPs over the time interval that started 50 ms after the beginning of the current fixation and ended at the median saccade initiation time (~215 ms) for the upcoming saccade. Neuronal responses were greater when the preferred stimulus was in the RF, across all conditions (Fig. 2, A and B) (*t* test; color, $P < 0.005$; shape, $P < 10^{-4}$). More importantly, the response to the preferred stimulus in the RF was enhanced when it also happened to match the cue, i.e., when it contained the target feature that the animal was searching for on that trial but had not yet found (color, $P < 10^{-5}$; shape, $P = 10^{-5}$). The response to the cued feature in the RF remained elevated through the time of saccade initiation, when spatially directed attention to the saccade target outside the RF was presumably at a maximum. Responses to nonpreferred RF stimuli showed no such enhancement based on the cue features ($P > 0.05$), but responses to distracters with colors similar to the preferred color were significantly enhanced when the preferred color was cued during color search (not shown, $P < 0.001$). The distribution of cue-feature effects on responses to a preferred stimulus in the RF is shown for all neurons in Fig. 2, G and H. More than 90% of the neurons gave a larger response to the preferred feature when it was the target feature (chi-square, $P < 10^{-5}$) with a median increase in response of 30% for all neurons. This supports the idea of an attentional bias in favor of neurons with a feature preference that matches the searched-for feature on a given trial, in parallel throughout the visual field.

In contrast to the effects of attention on single-neuron firing rates, neither the magnitude (Fig. 2, C and D) nor the spectral power of the LFP considered by itself were affected by the features of the stimulus in the RF or the searched-for cue features. However, as shown in the population data in Fig. 2, E and F, the coherence between spikes and the LFP in the gamma band (averaged over the 30- to 60-Hz range) was greater when the RF stimulus was of the preferred feature for the neurons compared to a nonpreferred feature (color, $P < 10^{-5}$; shape, $P = 0.001$). Most importantly, the coherence for the preferred feature in the RF was enhanced when the RF contained the target feature that the animal was searching for on that trial but had not yet found (cue effect: color, $P < 10^{-5}$; shape, $P < 10^{-5}$), similar to the effects we found on firing rates. More than 79% of the spike-LFP pairs showed enhanced coherence under these conditions (chi-square, $P < 10^{-5}$) (Fig. 2, G and H), with a median increase in coherence of 22% for all pairs (*25*). This supports the idea of a parallel bias in favor of neurons that prefer the feature of the searched-for target and happen to hold the undetected target within their RFs. Coherence for distracters with colors similar to the preferred color was also significantly enhanced when the preferred color was cued during color search (not shown, $P < 0.001$), suggesting that the bias in favor of the target feature was shared to a lesser extent by similar nontarget features. Finally, unlike what we found with firing rates, we found a small enhancement of coherence for a nonpreferred stimulus in the RF when the animal was searching for a feature that was the preferred feature of the neuron (cue effect: color, $P = 0.001$; shape, $P = 0.002$). For example, if the animals were searching for red, then red-preferring

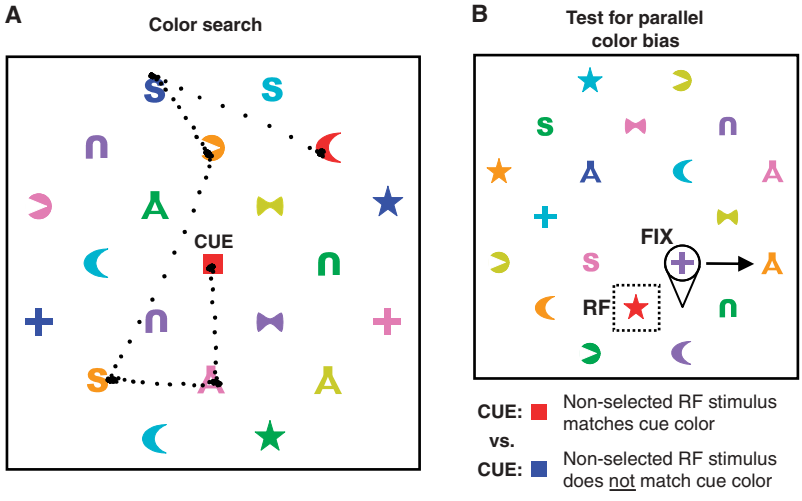


Fig. 1. Illustration of the color search task and the feature enhancement analysis. (**A**) An example of a color search trial. The cue at the center of the screen is shown for illustration purposes only; it was extinguished before the array onset in the experiment [as in (B)]. The black dots show the eye position of the monkey during a representative correctly performed trial. The colors and shapes of stimuli at each location changed pseudorandomly from trial to trial within a session (*20*). (**B**) Feature enhancement analysis during color search. We analyzed neuronal measures for stimuli in neurons' RF when those stimuli were not the goal of the impending saccade. In this example display, fixating (represented by the inverted cone) the purple cross brings the red star into the neuron's RF. This stimulus is not selected for the next saccade, which is made instead to the orange A. In this case, we would examine the response to the red star in the RF from the time it came into the RF (i.e., when the purple cross was fixated) to when it went out of the RF (i.e., the beginning of the saccade to the orange A). The critical cases were those in which the RF stimulus was of the preferred (e.g., red) or nonpreferred (e.g., blue) color for the neuron, and we analyzed responses to these two types of stimuli when they were either the search target (e.g., the cue was red and a red stimulus was in the RF) or they were a distracter (e.g., the cue was blue and a red stimulus was in the RF). The analysis was conducted in the same manner across all fixations that brought a stimulus of interest into the RF when that stimulus was not selected for a saccadic eye movement. The same analysis was conducted during shape search trials (fig. S1), but taking into account shape preference instead of color preference.

neurons had somewhat enhanced coherence even for stimuli in their RF that did not share the target feature. Together, these results suggest that when an animal is searching for a particular feature, the neurons that prefer that feature begin to synchronize their activity, and they go into maximum synchronization when a stimulus with that feature falls within their RF, e.g., when the animal is searching for red, the neurons prefer red, and a red stimulus falls within the RF.

Although the coherence measure we used was normalized for both firing rate and LFP power in a given frequency band, we confirmed that any overall changes in these values did not contribute to changes in coherence by calculating the coherence when the correspondence between spike and LFPs was randomly shuffled across fixations within a given condition (i.e., shuffling across fixations with the same cue and RF stimulus conditions). The shuffling not only reduced overall coherence but also eliminated the enhancement in the gamma band found for the preferred feature and preferred cue (Fig. 2, E and F).

Parallel selection during conjunction search. Another key element of most parallel models of visual search is that the top-down bias in favor of the cue-target stimulus representation in the cortex is shared by distracter stimuli if they share features with the cue target. For example, if the subject is searching for a horizontal red bar, the cortical representations of both horizontal stimuli and red stimuli will share in the target bias, including the cortical representation of vertical red distracters. If so, we should see evidence for enhancement of responses and/or synchrony when the stimulus in the RF contains a single feature of the target but is not itself a target because it lacks other target features. To test this, we studied 23 neurons (*26*) in one monkey during a conjunction task in which the target was defined by two features, color and shape, and the distracter elements could share only one of the target features; e.g., the target might be a red X, and a distracter might be either red, an X, or neither but would not contain both target features (fig. S1). We analyzed responses and synchronization associated with the distracters in the RF under the different cue conditions.

The monkey performed well (94% target localization) and did not search the array randomly (fig. S2A) (4.7 fixations per trial versus random, $P < 10^{-5}$; fixation duration versus required fixation duration for target detection, $P < 10^{-5}$). Behavioral data indicate that the target did not pop out (average of 90 ms per item to find the target in separate studies with a variable number of display items). As in the feature search task described above, and as in previous reports (*27*), the monkey used both the color and shape of the cue to guide its behavior (*28*).

Neuronal responses, LFPs, and spike-field synchrony were measured under four conditions: (i) the RF distracter contained the neuron's preferred feature and this feature was one of the cue (and target) attributes; (ii) the RF distracter contained the preferred feature but this feature was not cued; (iii) the RF distracter contained the nonpreferred feature and this feature was cued; and (iv) the RF distracter contained the nonpreferred feature when

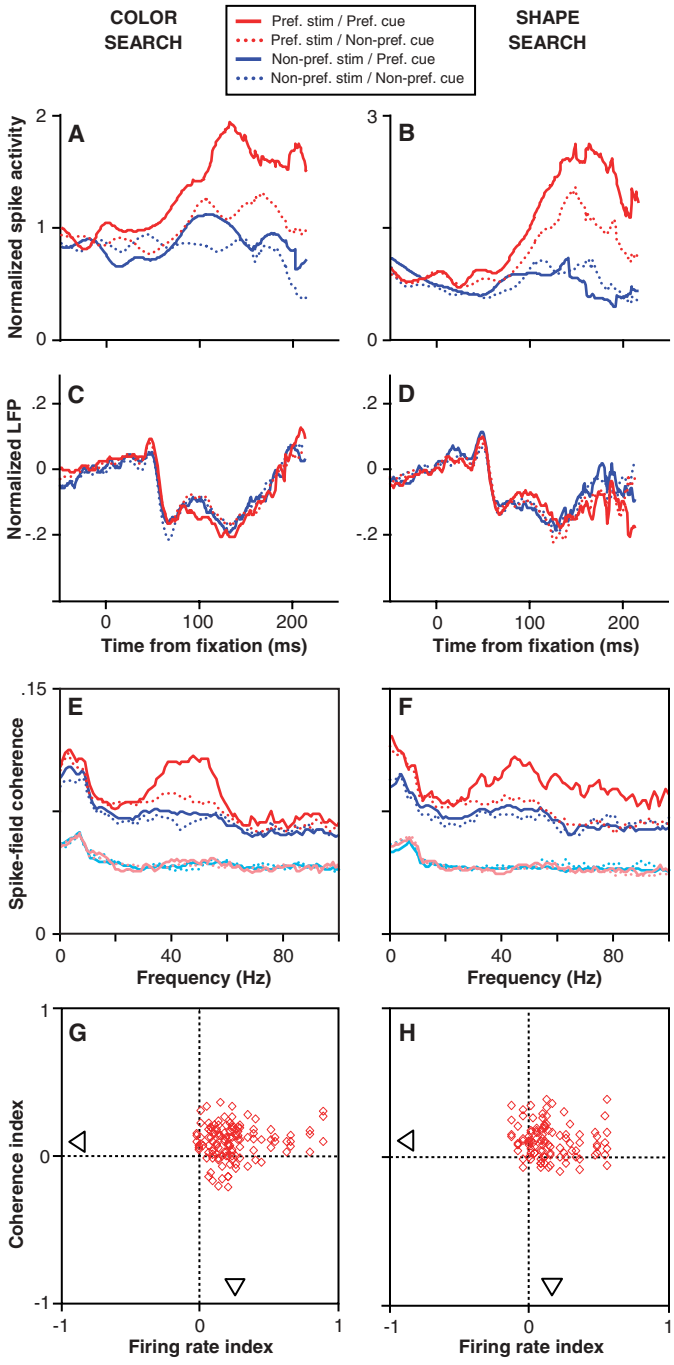


Fig. 2. Feature-related enhancement of neuronal activity and spike-field synchronization during feature searches. (**A**) Normalized firing rates averaged over a population of V4 neurons during color search trials during fixations at the end of which the monkey made a saccade away from the RF, averaged across all recordings. Red lines show responses when the stimulus in the RF was of the preferred color for the recorded neurons; blue lines show responses when the stimulus was of the nonpreferred color; solid lines show responses on trials in which the cue was the preferred color; and dotted lines show responses on trials in which the cue was the nonpreferred color. (**B**) Results for the shape search task, as in (A) but taking into account shape preference instead of color preference. (**C** and **D**) Normalized V4 LFP during color and shape search, respectively. (**E** and **F**) Spike-field coherence for the color and shape tasks, respectively. Light red and light blue lines show coherence values for the same conditions represented by the red and blue lines, respectively, but with the correspondence between LFP and single-neuron activity removed by shuffling the fixation order of the spike data within a condition. (**G** and **H**) The coherence modulation index as a function of the single-neuron firing rate modulation index for color and shape search, respectively, for each spike-LFP pair. The coherence modulation index was calculated as [(Cpp − Cpn)/(Cpp + Cpn)] and the firing rate modulation index as [(FRpp − FRpn)/(FRpp + FRpn)], where Cpp represents coherence when the preferred feature was in the RF and the preferred feature was cued, Cpn represents coherence when the preferred feature was in the RF and the nonpreferred feature was cued, and FRpp and FRpn represent firing rates for the same conditions as Cpp and Cpn, respectively. Positive indices reflect an enhancement of response to the preferred stimulus in the RF when the cue was also the preferred feature. Firing rate was averaged from 50 ms after fixation to the median saccade initiation time. Coherence was averaged over the 30- to 60-Hz range. The arrowheads indicate the mean coherence and firing rate indices across all pairs.

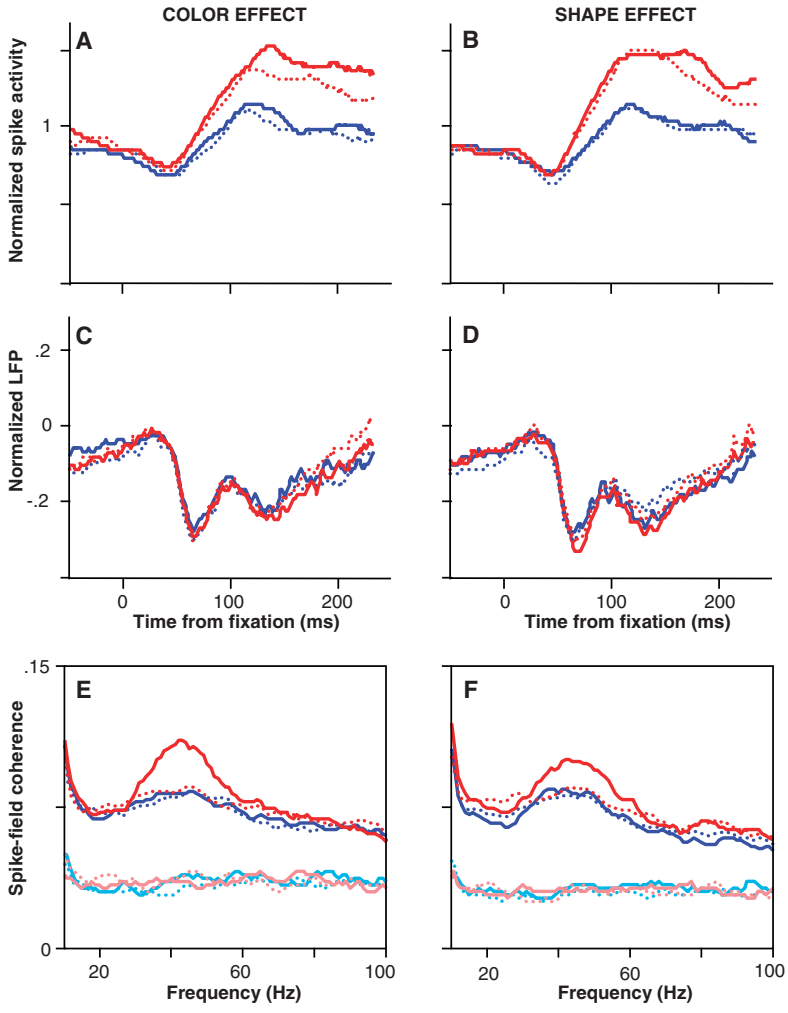


Fig. 3. (A to F) Feature-related enhancement of neuronal activity and synchronization during conjunction search. Conventions are as given in Fig. 2.

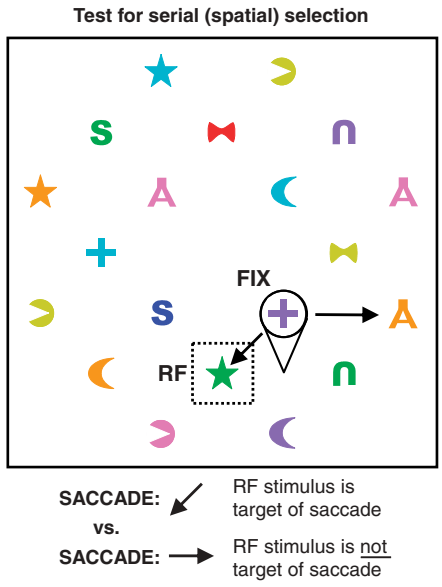


Fig. 4. Illustration of the saccade enhancement analysis. We compared neuronal measures when the monkey made a saccade to an RF stimulus versus a saccade away from the RF. In this display, fixating the purple cross, for example, brings the green star into the neuron's RF. We would then compare neuronal responses when the green star in the RF was the target of the saccade, to those when the saccade target was to a stimulus outside the RF, e.g., the orange A. Activity was analyzed from the time the purple cross was fixated to when the next saccade was initiated.

the preferred feature was cued (*22*). Neurons responded better to their preferred feature in the RF compared to nonpreferred features (Fig. 3, A and B) (color, $P < 0.01$; shape, $P < 0.001$). In the key test, we found that responses were enhanced if the distracter in the RF was of the neuron's preferred color and it was also the same color (but, by design, not the same shape) as the color-shape conjunction target (Fig. 3A) ($P = 0.002$). In other words, the distracter shared in the bias for the target stimulus if it shared one of its features, consistent with the predictions of parallel search models. The median enhancement was 8%, with more than 86% of the neurons having a larger response when the RF stimulus shared a feature with the searched-for target (chi-square, $P < 0.005$). There was also an enhancement of the response when the shape of the distracter matched the shape of the color-shape conjunction target, consistent with parallel models, but this enhancement was smaller and developed later than the color-related enhancement (Fig. 3, A and B). When the RF distracter was of the preferred feature, shape-related enhancement was not significant in the same time interval as that used in the feature search task, but it became significant ~150 ms after fixation onset ($P = 0.035$). This is consistent with the behavioral evidence described above, that the monkey used the color feature more than the shape feature in guiding its search to the color-shape conjunction target (fig. S2B). The LFP magnitude (Fig. 3, C and D) and power were not modulated by stimulus or cue features in the conjunction task.

There was also significant enhancement of the spike-field coherence in the gamma band when the RF distracter had the neuron's preferred feature and that feature was in common with the target for either a color (Fig. 3E) ($P < 10^{-5}$) or shape (Fig. 3F) ($P < 0.001$) match. The enhancement in the latter case was smaller, again consistent with the monkey's behavioral bias in favor of using color information. The median enhancement of coherence with a color match was 22%, with 97% of spike-LFP pairs showing an increase (chi-square, $P < 10^{-5}$), and the median enhancement with a shape match was 17%, with 78% of spike-LFP pairs showing an increase ($P < 0.002$). Thus, the top-down bias in visual search is not limited to cases in which the RF stimulus is the search target but instead applies to any stimulus, even a distracter, that contains a feature relevant to the search, consistent with parallel models. It is also consistent with the results from the feature search task, in which we found that enhancement occurred for colors that were similar to the target color. Both results potentially explain why search is often more difficult when the distracters share features with the target, as in some forms of conjunction search (*8*).

Serial selection during search. Finally, although we have emphasized the evidence for parallel mechanisms in search, the task necessarily had a spatial attention (serial) component to it, in that the animals made several saccades to stimuli in the array while searching for the targets. To test for spatial attention effects on responses, we compared responses and spike-field synchronization to a stimulus in the RF when either it was selected for a saccade or the saccade was made to a stimulus outside the RF (Fig. 4).

Selecting the RF stimulus for a saccade led to an enhancement of the neuronal response across the population (Fig. 5A) (population median enhancement of 36%, $P < 10^{-5}$,

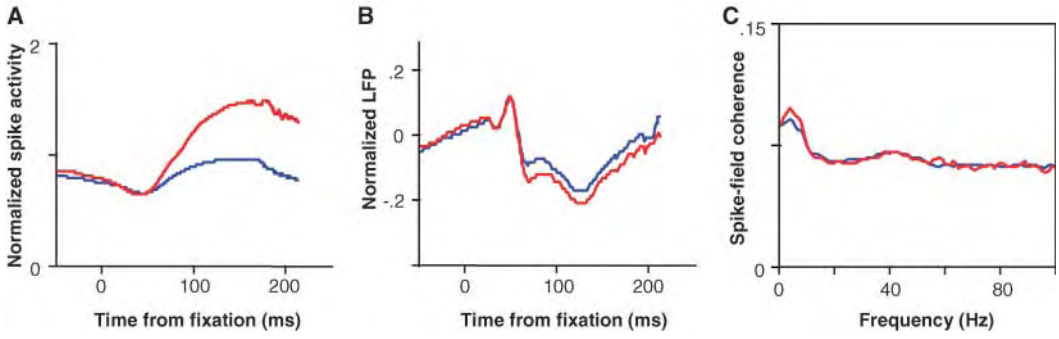


Fig. 5. Saccadic enhancement during feature searches. (**A**) Normalized firing rates for the population of neurons when a saccade was made to a stimulus inside the RF (red line) and when a saccade was made to a stimulus outside the RF (blue line) across all saccades. Data from color and shape searches were combined. (**B** and **C**) Normalized LFP and spike-field coherence for the same conditions. The results during conjunction search (not shown) were very similar.

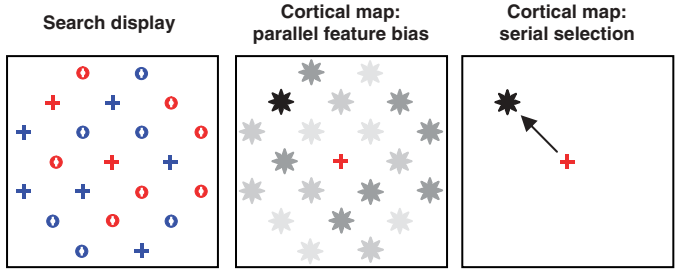


Fig. 6. Schematic illustration of selection mechanisms during a conjunction search task. (**Left**) Example stimulus display. (**Middle**) Representation of stimulus display in the cortex. Neurons that prefer the features of the searched-for target show enhanced firing and synchronize their activity when a stimulus sharing a target feature falls within their RFs, illustrated by the intensity of the stimulus representation. The greater the similarity or shared features with the target, the stronger the signal. This bias in favor of a potential target occurs in parallel throughout the visual field. (**Right**) The bias for potential targets triggers spatial selection mechanisms that result in eye movements.

with 70% of neurons showing a significant increase), and it also caused a significant modulation of the magnitude of the LFP (Fig. 5B) (population, $P < 10^{-5}$) and an increase in its spectral power in the gamma frequency range ($P < 0.001$). The effects were qualitatively similar if we aligned neuronal responses to the time of the saccade initiation and are consistent with previous reports (*10*).

In contrast to the parallel biasing effects during feature search on spike-field coherence, we found no significant effects on coherence of making a saccade to the RF stimulus (Fig. 5C) (*29*). Previously, we found that spatial attention enhances synchrony in V4, in a spatial attention task in which the animal was required to sustain attention to the same stimulus for up to several seconds (*17*). Given the much shorter stimulation intervals (~215-ms saccade dwell time) in the present study, it is possible that we were simply unable to detect a small change in coherence with spatial attention because there were too few spikes. The fact that the gamma-band power of the LFP was enhanced with attention during this initial interval suggests that the inputs to and/or activity within V4 were, in fact, becoming synchronized during this brief period, even though it was not yet evident in the spike-field coherence. In addition, recent work in our lab suggests that spike-field coherence builds over the course of sustained attention to a stimulus (*30*), which may also explain why increases in feature-related coherence are so prominent in visual search, where attention to the features of the searched-for object must be sustained for long intervals, even when individual objects in the scene are attended only briefly.

Discussion and conclusions. The results from the feature-selection and spatial-selection components of the search task together reveal the interplay of both parallel and serial neural mechanisms in visual search. As predicted by parallel-search models, including biased competition (*7*, *31*), the search for a target with a specific feature appears to synchronize and enhance the activity of the population of V4 neurons that prefer that feature, in parallel throughout the visual field representation, long before the animal locates the target. When a stimulus with a feature shared with the searched-for target then falls within the RF of the neurons preferring that feature, neural responses and synchrony are maximally enhanced (Fig. 6). Thus, the structures involved in the attentional control of area V4 must influence neurons based on their feature preferences and not just their RF locations. The V4 neurons firing synchronously, at an increased rate, will be more effective in driving postsynaptic neurons, including in inferotemporal and frontal cortex. We propose that this strong signal is the one that ultimately triggers spatial attention to the candidate target, and, in most cases, an eye movement toward it (Fig. 6). Distracter stimuli that resemble the target or that share some, but not all, features with the searched-for target also appear to share in this bias, explaining why some visual search tasks are difficult, including some where targets are defined by the conjunction of different features (*8*). This account, supported by studies of attention to multiple stimuli in V4 and inferotemporal RFs (*32*), is quite different from ideas that spatially focused attention is needed to "bind" together features such as color and shape in conjunction arrays or that search arrays are scanned by an internal attentional "spotlight" until the target is found. Nonetheless, there are important serial components to visual search. While the animals are searching, they appear to use a spatial-selection mechanism to examine some stimuli as potential targets, and the stimuli so selected elicit maximal neuronal responses. One can easily imagine that serial and parallel mechanisms are engaged to various degrees depending on the difficulty of the task and the sharing of target features among distracters, as predicted by hybrid models such as guided search (*8*) and FeatureGate (*33*).

Overall, it appears that processing in V4 is an intermediate stage in visual search between stimulus-feature processing and high-level object recognition. The feature-related enhancement we observed is likely the result of a combination of feature-selective responses in the visual cortex, including V4, and top-down feedback from structures involved in working memory and executive control, such as the prefrontal cortex and possibly the parietal cortex (*7*, *34*). Such feedback must be capable of targeting neurons with the appropriate feature preferences throughout the visual field map. The saccade-related enhancement, on the other hand, likely originates from feedback to V4 neurons with RFs at particular locations, originating from structures with spatial attention and oculomotor functions such as the frontal eye field and the lateral intraparietal area. These areas are thought to represent a salience map in which stimuli are represented according to their behavioral relevance independent of their features (*35*), ultimately resulting in the selection of a single stimulus for a saccade target or further visual processing (*36*).

References and Notes

1. S. P. Liversedge, J. M. Findlay, *Trends Cogn. Sci.* **4**, 6 (2000).
2. S. J. Luck, L. Chelazzi, S. A. Hillyard, R. Desimone, *J. Neurophysiol.* **77**, 24 (1997).

3. E. Seidemann, W. T. Newsome, *J. Neurophysiol.* **81**, 1783 (1999).
4. C. L. Colby, M. E. Goldberg, *Annu. Rev. Neurosci.* **22**, 319 (1999).
5. S. Treue, J. H. R. Maunsell, *J. Neurosci.* **19**, 7591 (1999).
6. R. A. Andersen, C. A. Buneo, *Annu. Rev. Neurosci.* **25**, 189 (2002).
7. R. Desimone, J. Duncan, *Annu. Rev. Neurosci.* **18**, 193 (1995).
8. J. M. Wolfe, K. R. Cave, S. L. Franzel, *J. Exp. Psychol. Hum. Percept. Perform.* **15**, 419 (1989).
9. L. Chelazzi, E. K. Miller, J. Duncan, R. Desimone, *Cereb. Cortex* **11**, 761 (2001).
10. J. A. Mazer, J. L. Gallant, *Neuron* **40**, 1241 (2003).
11. T. Ogawa, H. Komatsu, *J. Neurosci.* **24**, 6371 (2004).
12. P. E. Haenny, J. H. Maunsell, P. H. Schiller, *Exp. Brain Res.* **69**, 245 (1988).
13. B. C. Motter, *J. Neurosci.* **14**, 2178 (1994).
14. C. J. McAdams, J. H. R. Maunsell, *J. Neurophysiol.* **83**, 1751 (2000).
15. P. Fries, P. R. Roelfsema, A. K. Engel, P. Konig, W. Singer, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12699 (1997).
16. M. R. Jarvis, P. P. Mitra, *Neural Comput.* **13**, 717 (2001).
17. P. Fries, J. H. Reynolds, A. E. Rorie, R. Desimone, *Science* **291**, 1560 (2001).
18. B. Pesaran, J. S. Pezaris, M. Sahani, P. P. Mitra, R. A. Andersen, *Nat. Neurosci.* **5**, 805 (2002).
19. E. Salinas, T. J. Sejnowski, *J. Neurosci.* **20**, 6193 (2000).
20. Materials and methods are available as supporting material on *Science* Online.
21. Several lines of evidence suggest that the monkeys used the feature of the cue to guide their search and did not make fixations around the array randomly. First, we did not find any evidence for regular scan patterns, and the average number of saccades to find the target under the different conditions was always significantly less than the number of saccades that would be required to find the target by chance (fig. S2A) (*t* tests, $P < 10^{-5}$). Second, the median fixation duration across those saccades was less than half and significantly less than the minimum 700-ms fixation duration required to correctly signal the acquisition of the target (fig. S2A) ($P < 10^{-5}$). Finally, the pattern of fixations on distracters indicated that the target color had a large effect on the distracters that were fixated. During color search, monkeys were significantly more likely to fixate a distracter with a color similar to the cued target color (e.g., an orange or magenta stimulus if the cue was red) or distracters with the second cued color in a given session than they were to fixate other distracters (fig. S2B) (analysis of variance, $P < 10^{-5}$; group comparisons, $P < 0.05$). Likewise, they also fixated distracters with colors similar to the cued color significantly longer than other distracters (fig. S2B) (analysis of variance, $P < 10^{-5}$; group comparisons, $P < 0.05$). For shape search, there was no reliable method to measure perceptual similarity among shapes. However, consistent with the results in color search, the monkeys showed significant differences in fixations on distracters depending on the shape of the searched-for target.
22. Across the population, there were no significant differences in either the average fixation number or the relative contribution of the first fixation after array onset to the data used in comparing the four conditions of interest for each visual search.
23. Fifteen sessions were performed with one monkey and twelve with the other.
24. Forty-four neurons (in 23 sessions) contributed to the analysis of feature enhancement during color search based on significant color selectivity for the two color cues used, and 38 neurons (in 21 sessions) contributed to the analysis of feature enhancement during shape search based on significant shape selectivity. Sixty-one and 57 LFP signals contributed to the analyses of color and shape search, respectively.
25. We calculated the coherence between spikes on one electrode and the LFP recorded on all other electrodes separately. It made no difference in the overall results whether the LFP from the same electrode was used or not, and therefore we included it in our analyses.
26. Of these neurons, 15 contributed to the analysis of color enhancement and 16 to the analysis of shape enhancement.
27. N. P. Bichot, J. D. Schall, *Nat. Neurosci.* **2**, 549 (1999).
28. Fixation probability was greater and fixation duration increased for distracters that shared one of the cue features, although there was a clear bias in favor of selecting stimuli that matched the color of the target compared to stimuli matching the shape (fig. S2B) (analysis of variance, $P < 10^{-5}$; group comparisons, $P < 0.05$).
29. Using the conjunction search data, we found that the lack of saccadic enhancement in the spike-field coherence was not dependent on the properties of the RF stimulus.
30. E. A. Buffalo, R. Desimone, unpublished data.
31. J. H. Reynolds, L. Chelazzi, R. Desimone, *J. Neurosci.* **19**, 1736 (1999).
32. J. Moran, R. Desimone, *Science* **229**, 782 (1985).
33. K. R. Cave, M. S. Kim, N. P. Bichot, K. V. Sobel, in *Neurobiology of Attention*, L. Itti, G. Rees, J. Tsotsos, Eds. (Elsevier, Boston, 2004), pp. 547–552.
34. E. K. Miller, C. A. Erickson, R. Desimone, *J. Neurosci.* **16**, 5154 (1996).
35. K. G. Thompson, N. P. Bichot, *Prog. Brain Res.* **147**, 251 (2005).
36. F. H. Hamker, *Cereb. Cortex* **15**, 431 (2005).

37. We thank Plexon, Inc., for technical assistance, and P. Fries, P. Mitra, K. Thompson, and J. Schall for valuable comments. Supported by the NIMH Intramural Research Program.

Supporting Online Material
www.sciencemag.org/cgi/content/full/308/5721/529/DC1
Materials and Methods
Figs. S1 and S2
References and Notes

11 January 2005; accepted 14 March 2005
10.1126/science.1109676

REPORTS

Sub–Diffraction-Limited Optical Imaging with a Silver Superlens

Nicholas Fang, Hyesog Lee, Cheng Sun, Xiang Zhang*

Recent theory has predicted a superlens that is capable of producing sub–diffraction-limited images. This superlens would allow the recovery of evanescent waves in an image via the excitation of surface plasmons. Using silver as a natural optical superlens, we demonstrated sub–diffraction-limited imaging with 60-nanometer half-pitch resolution, or one-sixth of the illumination wavelength. By proper design of the working wavelength and the thickness of silver that allows access to a broad spectrum of subwavelength features, we also showed that arbitrary nanostructures can be imaged with good fidelity. The optical superlens promises exciting avenues to nanoscale optical imaging and ultrasmall optoelectronic devices.

Conventional optics suffer from the so-called diffraction limit, because they are only capable of transmitting the propagating components emanating from the source (*1*). This is because the evanescent waves that carry subwavelength information about the object decay exponentially in a medium with positive permittivity and permeability and are lost before reaching the image plane. To avoid this evanescent decay, a concept of subwavelength imaging through an intimate-contact mask was proposed back in the 1970s (*2*) and was subsequently demonstrated (*3*). Conventional immersion lenses that improve image resolution by increasing the index of refraction are limited by the availability of high-index materials. Although scanning near-field optical microscopy detects subwavelength details, it uses point-by-point scanning rather than forming the whole image at once. Interference can also form simple subwavelength patterns (*4*, *5*), but it is not considered to be an imaging method because of the lack of a general bijective conformal mapping between the object and the image.

In contrast, a superlens is predicted to substantially enhance evanescent waves (*1*), compensating for the evanescent loss outside the superlens and thus restoring an image below the diffraction limit. This unusual lens takes the form of a thin slab of material with negative permittivity or permeability, or both (*1*, *6*–*10*). Metamaterials (*6*–*15*), an emerging class of artificially designed and structured materials, have shown unprecedented electromagnetic properties in microwave or terahertz regimes, which provide a basis for the conjecture of the superlens. Although metamaterials and photonic crystals have recently shown beam bending (*8*, *13*, *14*) and refocus-

5130 Etcheverry Hall, Nanoscale Science and Engineering Center, University of California, Berkeley, CA 94720–1740, USA.

*To whom correspondence should be addressed. E-mail: xiang@berkeley.edu

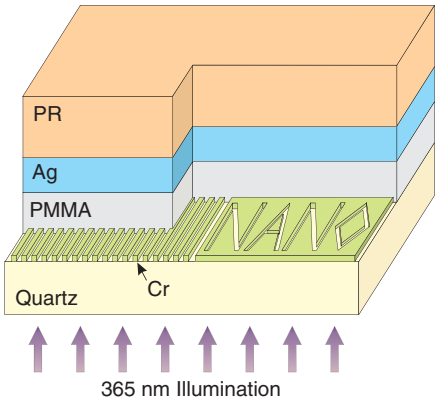


Fig. 1. Optical superlensing experiment. The embedded objects are inscribed onto the 50-nm-thick chrome (Cr); at left is an array of 60-nm-wide slots of 120 nm pitch, separated from the 35-nm-thick silver film by a 40-nm PMMA spacer layer. The image of the object is recorded by the photoresist on the other side of the silver superlens.

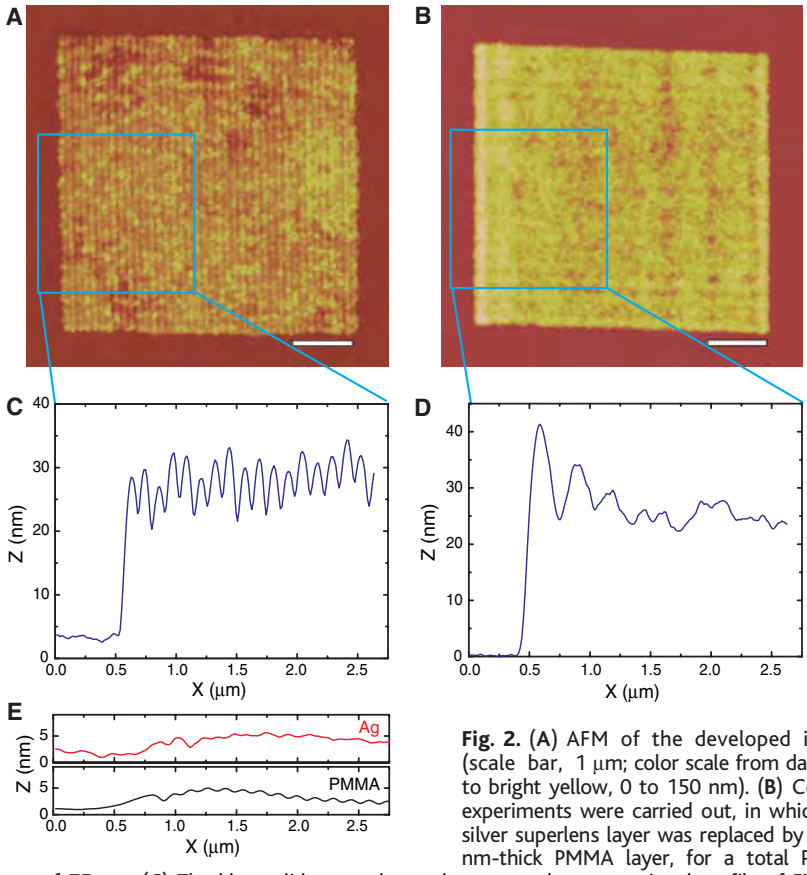


Fig. 2. (**A**) AFM of the developed image (scale bar, 1 μm; color scale from dark red to bright yellow, 0 to 150 nm). (**B**) Control experiments were carried out, in which the silver superlens layer was replaced by a 35-nm-thick PMMA layer, for a total PMMA thickness of 75 nm. (**C**) The blue solid curve shows the averaged cross-sectional profile of Fig. 2A, clearly demonstrating the 63 ± 4 nm half-pitch resolved with a 35-nm silver superlens. *X*, relative displacement along the cross-section direction. (**D**) The blue solid curve shows the average cross section of Fig. 2B (control sample). (**E**) The red and black solid curves show the corresponding surface roughness profile on silver and the PMMA layer above the grating object, respectively.

ing (*6*, *7*, *9*, *10*, *16*) at microwave frequencies, achieving superlensing at optical frequencies is challenging partly because of the diminishing magnetic susceptibility of natural materials. Only recently have metamaterials shown magnetic responses in terahertz and infrared regimes (*11*, *12*). However, in the near field, the electric and magnetic responses of materials are decoupled; therefore, for transverse magnetic (TM) waves, only the permittivity needs to be considered (*1*). This makes noble metals such as silver natural candidates for optical superlensing, because negative permittivity is easily attainable in them as a consequence of the collective excitation of conduction electrons. Let us consider a superlens consisting of a silver slab separated from an object by a spacer layer, and coated on the opposite side with an imaging material. By designing the thin metal slab so that the surface current oscillations (the surface plasmons) match the evanescent waves from the object, the superlens is able to substantially enhance the amplitude of the field. This key precursor for superlensing—the enhancement of evanescent waves by surface plasmons—was observed on silver films (*17*). Recently, numerical simulations have been conducted to understand the details of the silver superlens (*18*–*22*). However, experimental attempts reported diffraction-limited images, probably resulting from the use of thick silver films that hinder the enhancement of evanescent waves (*20*).

As surface charges accumulate at the interface between the silver and the imaging medium, the normal component of an electric field is found at resonance when a proper thickness of silver is selected and the permittivity of the silver and that of the adjacent medium are equal and of opposite sign. Such a delicate resonance is essential to ensure the evanescent enhancement across the slab (*1*, *22*). For enhanced transmission of evanescent waves (*23*), it is found that an asymptotic impedance match ($k_{zi}/\varepsilon_i + k_{zj}/\varepsilon_j = 0$) has to be met at the surface of the silver, known as the surface plasmon excitation condition (k_{zi}, cross-plane wave vector in silver; ε_i, permittivity of silver; k_{zj}, cross-plane wave vector in dialectric; and ε_j, permittivity of dialectric). It is widely known in metal optics that when the two media take the opposite sign in permittivity and $|\varepsilon_i| >> \varepsilon_j$, only surface plasmons at the narrow range of in-plane wave vector (k_x) that are close to k_0 can be resonantly coupled. However, less well known is that when $|\varepsilon_i| \sim \varepsilon_j$ and they are of opposite sign, the excitable surface plasmon band of k_x is significantly broadened, resulting in the superlensing effect (*1*).

Figure 1 shows a set of objects inscribed onto a chrome screen. The objects, placed about 40 nm away from the 35-nm silver film, are imaged onto the photoresist on the other side of the silver film under ultraviolet (UV) illumination (at a wavelength of 365 nm). Using focused ion beam (FIB) lithography, the chrome objects are patterned on quartz and then planarized, using a 40-nm-thick layer of polymethyl methacrylate (PMMA). The chrome masks are approximated as a binary object because in the UV, chromium (permittivity = –8.55 + i8.96) is away from the plasmon resonance and possesses low skin depth (~15 nm) (*24*). Subsequently, a 35-nm-thick layer of silver, the superlens, is evaporated over the PMMA spacer layer, followed by coating with a 120-nm-thick negative photoresist [NFR105G, Japan Synthetic Rubber Microelectronics (JSR Micro)] to record the near-field image. The substrate is then flood-exposed under an I-line (365-nm) mercury lamp (*25*). The exposure flux is 8 mW/cm^2, and an optimal exposure time of 60 s is applied. It is essential to reduce the surface root mean square modulation below 1 nm for both the silver and PMMA surfaces (Fig. 2E); otherwise, the dispersion characteristics of the superlens would be modified and would in turn smear the details of the recorded image (*19*). The optical image is converted into topographic modulations by developing the negative photoresist and is mapped using atomic force microscopy (AFM). The image of an array of 60-nm-wide nanowires on a 120-nm pitch is shown in Fig. 2A. The Fourier-transformed spectrum shows a sharp peak, corresponding to a pitch of 126 ± 7 nm (Fig. 3A, bottom panel). With careful control

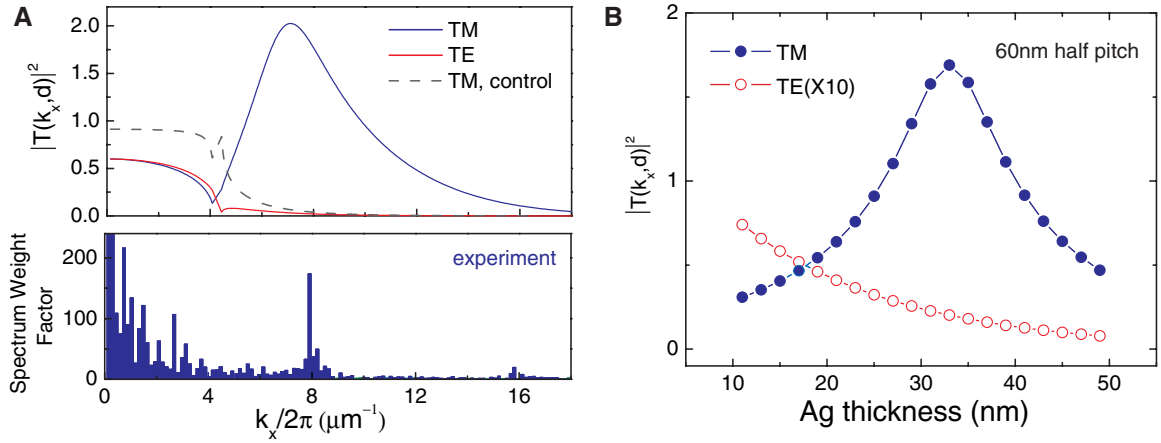


Fig. 3. The computed transfer function of the superlens system (the square modulus of the image, *E*-field normalized by object *E* field in wave vector space) (*27*). (**A**) (Top) Blue and red solid curves show the overall transmission of TM and TE waves over the combination of a 40-nm PMMA spacer layer and a 35-nm silver film, respectively; the black dashed curve is calculated by replacing the 35-nm silver layer with a 35-nm PMMA layer (TM polarization) in the control experiment. The abscissa is the normalized lateral wave vector. *T*, transmission coefficient; *d*, thickness of silver. (Bottom) The navy columns show the experimentally resolved spectral components of the imaged object. The major peak indicates a 63 ± 4 nm half-pitch. (**B**) The image transfer function for TM and TE fields versus silver thickness for 60-nm half-pitch objects.

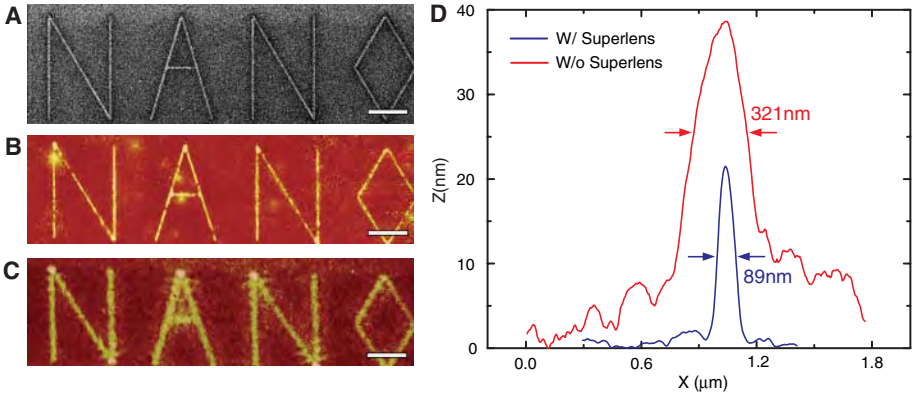


Fig. 4. An arbitrary object "NANO" was imaged by silver superlens. (**A**) FIB image of the object. The linewidth of the "NANO" object was 40 nm. Scale bar in (A) to (C), 2 μm. (**B**) AFM of the developed image on photoresist with a silver superlens. (**C**) AFM of the developed image on photoresist when the 35-nm-thick layer of silver was replaced by PMMA spacer as a control experiment. (**D**) The averaged cross section of letter "A" shows an exposed line width of 89 nm (blue line), whereas in the control experiment, we measured a diffraction-limited full width at half-maximum line width of 321 ± 10 nm (red line).

of the surface morphology of the PMMA and silver surface, we recorded the photoresist images with typical average height modulations of 5 to 10 nm (Fig. 2C). The result indicates sub–diffraction-limited imaging with half-pitch resolution down to 60 nm, or λ/6, using a silver superlens. The control experiments in which the silver film is replaced by 35-nm-thick PMMA show no evidence of imaging contrast from the subwavelength object, regardless of the optimization of imaging processes (Fig. 2, B and D). Our observation is further verified by a simple analysis of the evanescent decay of these subwavelength features. For a line-array object of period *a*, the intensity of evanescent waves decays with a characteristic length *Z*, given by $Z^{-1} = 4\pi\sqrt{a^{-2} - \varepsilon\lambda^{-2}}$ (*26*). With the permittivity of the surrounding media $\varepsilon \sim 2.4$, a decay length of 11 nm is estimated for the 60-nm half-pitch. Without a superlens to enhance and transmit evanescent waves, it is difficult to resolve a 60-nm half-pitch object from a distance of 75 nm away, as we observed in the control experiments.

For the perfect lens originally proposed by Pendry, the entire Fourier spectrum of the image should be transferred uniformly without loss or distortion. However, a slab lens made of real materials with dispersive negative permittivity (such as silver) or permeability inevitably has losses and only asymptotically approaches a perfect lens at the quasi-static limit. Yet, such a lens system greatly enhances the essential portion of the evanescent spectrum, allowing subwavelength imaging resolution. We term such system a superlens, in comparison to an ideally perfect lens. Superlensing requires rather stringent conditions to maximize the enhancement over a broad band of evanescent waves. It is crucial to find a wavelength range where the wavelength-dependent permittivity of silver is nearly equal to and opposite that of the adjacent medium, which is the PMMA spacer and photoresist in this experiment. The resolving power of this system can be measured by the optical transfer function, defined as the ratio of image field to object field, $|E|_{img}/|E|_{obj}$, with given lateral wave vector k_x. Using the Fresnel formula of a stratified medium (*18*), we computed the optical transfer function of a silver superlens system, taking into consideration the absorption in silver and in the adjacent media (Fig. 3). From Fig. 3A, we estimate that a range of evanescent components from 2 to 4 k_0 can be efficiently enhanced and recovered by surface plasmon excitations, which is much broader than the conventional narrow resonance (typically <10% of k_0). The calculation shows that the slight absorptions from PMMA and the photoresist have only negligible effects on the imaging quality.

Although the illumination is normal to the plane of the object, the subwavelength object scatters the incident radiation in different directions. Two types of polarization for the scattered waves in reference to the plane of the silver superlens are classified: (i) waves with a magnetic field component parallel to the plane have TM polarization, and (ii) waves with an electric field parallel to the plane have transverse electric (TE) polarization. Whereas a broad band of TM evanescent waves are resonantly coupled over the 35-nm silver superlens to form a sub–diffraction-limited image, the TE fields are significantly damped out (Fig. 3A), which is consistent with theoretical estimations of the image resolution calculated against the loss of natural silver (*18*, *19*). Even though the PMMA and the photoresist slightly detune the surface plasmon resonance at the interfaces, the finite thickness of the silver slab ensures that they will be excited (*22*), allowing enhanced transfer of a broad evanescent spectrum. Because the 60-nm half-pitch object carries fundamental wave vectors that fall within the broad band of enhancement by the superlens (Fig. 3A, top panel), the corresponding spectrum peak is clearly observed in the recorded image, as seen in our experiment (Fig. 3A, bottom panel). In

Fig. 3B, we calculated the transfer function of the 60-nm half-pitch object for various silver film thicknesses. Assisted by surface plasmon excitation, a maximum enhancement of the selected subwavelength feature was reached for a 35-nm-thick silver slab. Above a thickness of 40 nm, the enhancement was damped by material absorption, destroying the superlensing effect. Such a critical thickness gives a measure for the efficient coupling of the evanescent modes at two interfaces of the silver film. The existence of this critical thickness is in good agreement with our previous finding (*15*).

The silver superlens can also image arbitrary nanostructures with sub–diffraction-limited resolution (Fig. 4). The recorded image "NANO" in Fig. 4B shows that we can faithfully reproduce the fine features from the mask (Fig. 4A) in all directions with good fidelity. As previously discussed, only the scattered TM evanescent waves from the object are coupled into the surface plasmon resonance of the silver film, and they become a primary component for restoring a sub–diffraction-limited image. For comparison, Fig. 4C displays the control experiment performed on the same mask, with "NANO" embedded in 75-nm planarized PMMA. With the same exposure condition, we observed the image of "NANO" with much wider lines, even with the extended development time (>1 min). Because the lines are a few micrometers apart, the subwavelength openings can be treated as isolated line sources, with a broad band of Fourier spectrum: The larger Fourier components decay strongly, and only the smaller Fourier components reach the imaging plane, resulting in a diffraction-limited image as shown in Fig. 4C. In contrast, with a silver superlens, we can resolve an average line width of 89 nm (Fig. 4D), which is much smaller than that of the diffraction-limited image. With natural and structured plasmons (*28–31*) in metals and phonon polaritons in semiconductors, a potential low-loss, high-resolution superlens is achievable.

References and Notes

1. J. B. Pendry, *Phys. Rev. Lett.* **85**, 3966 (2000).
2. H. I. Smith, *Proc. IEEE* **62**, 1361 (1974).
3. U. Ch. Fischer, H. P. Zingsheim, *J. Vac. Sci. Technol.* **19**, 881 (1981).
4. J. G. Goodberlet, H. Kavak, *Appl. Phys. Lett.* **81**, 1315 (2002).
5. X. Luo, T. Ishihara, *Appl. Phys. Lett.* **84**, 4780 (2004).
6. M. C. K. Wiltshire *et al.*, *Science* **291**, 849 (2001).
7. M. C. K. Wiltshire *et al.*, *Opt. Express* **11**, 709 (2003).
8. R. A. Shelby, D. R. Smith, S. Schultz, *Science* **292**, 77 (2001).
9. A. K. Iyer, P. C. Kremer, G. V. Eleftheriades, *Opt. Express* **11**, 696 (2003).
10. A. Grbic, G. V. Eleftheriades, *Phys. Rev. Lett.* **92**, 117403 (2004).
11. T.-J. Yen *et al.*, *Science* **303**, 1494 (2004).
12. S. Linden *et al.*, *Science* **306**, 3531 (2004).
13. C. G. Parazzoli *et al.*, *Phys. Rev. Lett.* **90**, 107401 (2003).
14. A. A. Houck, J. B. Brock, I. L. Chuang, *Phys. Rev. Lett.* **90**, 137401 (2003).
15. D. R. Smith, J. B. Pendry, M. C. K. Wiltshire, *Science* **305**, 788 (2004).
16. P. V. Parimi, W. T. Lu, P. Vodo, S. Sridhar, *Nature* **426**, 404 (2003).
17. Z. Liu, N. Fang, T.-J. Yen, X. Zhang, *Appl. Phys. Lett.* **83**, 5184 (2003).
18. N. Fang, X. Zhang, *Appl. Phys. Lett.* **82**, 161 (2003).
19. D. R. Smith *et al.*, *Appl. Phys. Lett.* **82**, 1506 (2003).
20. D. O. S. Melville, R. J. Blaikie, *J. Vac. Sci. Technol.* **B22**, 3470 (2004).
21. D. O. S. Melville, R. J. Blaikie, *J. Opt.* **A7**, S176 (2005).
22. S. A. Ramakrishna, D. Schurig, D. R. Smith, S. Schultz, J. B. Pendry, *J. Mod. Opt.* **49**, 1747 (2002).
23. N. Fang, Z. Liu, T.-J. Yen, X. Zhang, *Opt. Express* **11**, 682 (2003).
24. In our estimation, an ideal mask was assumed to have perfect binary near-field profiles, but the actual intensity may be influenced by the material and geometries of the subwavelength mask near the superlens. Thus, Fig. 3 should be regarded as a general guideline to the performance of the real system. A more accurate description can be given by full vector electromagnetic simulations of the system, such as in (*21*).
25. The image-recording conditions have been optimized to obtain the best possible image contrast in all experiments in this report. See the details of the control experiment in the supporting material on *Science* Online.
26. E. Loewen, E. Popov, *Diffraction Gratings and Applications* (Marcel Dekker, New York, 1997).
27. To compute the image transfer function presented in Fig. 3, the dielectric constant ($2.3013 + 0.0014i$) of PMMA is taken from the 495PMMA product datasheet (Microchem), the refractive index ($n = 1.6099$) of NFR105G is taken from JSR Micro, and a typical value of $\kappa = 0.003$ is taken into account for the absorption of the photoresist. In the computation, we assume that the photoresist occupies the top half of the space, whereas experimentally the actual photoresist layer thickness is 120 nm. The permittivity of silver ($-2.4012 + 0.2488i$) is taken from (*32*).
28. J. B. Pendry, A. J. Holden, W. J. Stewart, I. Youngs, *Phys. Rev. Lett.* **76**, 4773 (1996).
29. H. J. Lezec *et al.*, *Science* **297**, 820 (2002).
30. T. W. Ebbesen, H. J. Lezec, H. F. Ghaemi, T. Thio, P. A. Wolff, *Nature* **391**, 667 (1998).
31. J. B. Pendry, L. Martín-Moreno, F. J. Garcia-Vidal, *Science* **305**, 847 (2004).
32. P. B. Johnson, R. W. Christy, *Phys. Rev. B* **6**, 4370 (1972).

33. The authors acknowledge support from the Office of Naval Research (ONR)/Defense Advanced Research Projects Agency Multidisciplinary University Research Initiative (MURI) (ONR grant N00014-01-1-0803) and from NSF for the Center for Nanoscale Science and Engineering (grant DMI-0327077).

Supporting Online Material

www.sciencemag.org/cgi/content/full/308/5721/534/DC1
Materials and Methods
Figs. S1 to S5
Table S1

16 December 2004; accepted 14 March 2005
10.1126/science.1108759

Monodisperse Double Emulsions Generated from a Microcapillary Device

A. S. Utada,[1] E. Lorenceau,[1] D. R. Link,[1,2] P. D. Kaplan,[3] H. A. Stone,[1] D. A. Weitz[1,2]

Double emulsions are highly structured fluids consisting of emulsion drops that contain smaller droplets inside. Although double emulsions are potentially of commercial value, traditional fabrication by means of two emulsification steps leads to very ill-controlled structuring. Using a microcapillary device, we fabricated double emulsions that contained a single internal droplet in a core-shell geometry. We show that the droplet size can be quantitatively predicted from the flow profiles of the fluids. The double emulsions were used to generate encapsulation structures by manipulating the properties of the fluid that makes up the shell. The high degree of control afforded by this method and the completely separate fluid streams make this a flexible and promising technique.

Mixing two immiscible fluids produces an emulsion, which is defined as a dispersion of droplets of one fluid in a second fluid. Although they are not in equilibrium, emulsions can be metastable, with the droplets retaining their integrity for extended periods of time if their interface is stabilized by a surfactant. Emulsions play critical roles in many forms of processing and in coatings, foods, and cosmetics (*1*). One common use is to compartmentalize one fluid in a second, which is particularly important for packaging and stabilizing fluids and other active ingredients. Even more control and flexibility for encapsulation can be achieved through the use of double emulsions, which are emulsions with smaller droplets of a third fluid within the larger drops. The intermediate fluid adds an additional barrier that separates the innermost fluid from the outer fluid, or the continuous phase. This makes double emulsions highly desirable for applications in controlled release of substances (*2–4*); separation (*5*); and encapsulation of nutrients and flavors for food additives (*6–8*); and for the control of encapsulation, release, and rheology for personal care products (*4*, *9–11*). Additional flexibility is achieved by controlling the state of the

[1]Division of Engineering and Applied Sciences, [2]Department of Physics, Harvard University, Cambridge, MA 02138, USA. [3]Unilever Skin Global Innovation Center, 40 Merritt Boulevard, Trumbull, CT 06611, USA.

middle fluid, which can be selectively gelled or hardened to create solid capsules (*12*, *13*). These capsules can be used to encapsulate of drugs for targeted delivery and release (*14*–*21*).

Multiple emulsions are typically made in a two-step process, by first emulsifying the inner droplets in the middle fluid, and then undertaking a second emulsification step for the dispersion (*22*). Each emulsification step results in a highly polydisperse droplet distribution, exacerbating the polydispersity of the final double emulsion. Thus, any capsules formed from such double emulsions are, by nature, poorly controlled in both size and structure, and this limits their use in applications that require precise control and release of active materials. Microfluidic techniques can circumvent the vagaries of the bulk emulsification process and can produce more uniform double emulsions (*23*), although the range of drop sizes is limited and the devices require localized surface functionalization to control wettability in order to function. Alternatively, flow focusing of coaxial jets (*24*) can produce uniform coated droplets, but these must be re-emulsified into the continuous phase, which is a difficult step that precludes widespread use of this technique. The availability of highly monodisperse double emulsions would not only greatly improve their applicability but would also allow for detailed studies of their stability under more controlled conditions (*25*–*31*).

Here we describe a fluidic device that generates double emulsions in a single step, allowing precision control of the outer and inner drop sizes as well as the number of droplets encapsulated in each larger drop. Our device consists of cylindrical glass capillary tubes nested within a square glass tube. By ensuring that the outer diameter of the round tubes is the same as the inner dimension of the square tube, we achieve good alignment to form a coaxial geometry. The innermost fluid is pumped through a tapered cylindrical capillary tube, and the middle fluid is pumped through the outer coaxial region (Fig. 1A), which forms a coaxial flow at the exit of the tapered tube. The outermost fluid is pumped through the outer coaxial region from the opposite direction, and all fluids are forced through the exit orifice formed by the remaining inner tube (Fig. 1A). This geometry results in hydrodynamic focusing (*24*, *32*) of the coaxial flow. The flow passes through the exit orifice and subsequently ruptures to form drops; however, the coaxial flow can maintain its integrity and generate double emulsion droplets within the collection tube. We were also able to produce single emulsions by removing the tapered inner injection tube. In this geometry, our device is reminiscent of the selective withdrawal technique (*33*). Typical diameters of the exit orifice in our devices range from 20 to 200 μm; however, smaller or larger orifices can also be used, which allows the drop size to be adjusted. For convenience, we used a collection tube whose inner diameter was initially narrow and abruptly widened at a distance equal to one–article diameter downstream. These collection tubes were fabricated by axially heating the end of a cylindrical glass tube; as the glass liquefies, the orifice shrinks. Alternatively, we can use a tapered capillary tube as the collection tube. This can provide additional control, but the alignment of two tapered capillary tubes is more delicate and hence more difficult.

We achieved a high degree of control over the resultant double emulsions, varying the diameters of both the outer and inner drops and the number of inner droplets [supporting online material (SOM) text I and fig. S1]. We can produce uniform double emulsions, in which each drop contains a single internal droplet, creating core-shell structures whose drop diameter and shell thickness can be controlled. For example, we can form drops with extremely thin shells; the ratio of shell thickness to outer drop radius can be as low as 3% (Fig. 1B). Alternatively, we can increase the shell thickness up to about 40% of the drop radius (Fig. 1, C to E). We can also vary the number and size of the internal droplets in the double emulsions (Fig. 1, F and G). A stream of double emulsions, each containing a single internal droplet, is shown in Fig. 1H.

To gain insight into the breakup of a coaxial flow, we first considered the formation of single emulsions. We defined two mechanisms of drop formation for our device geometry: dripping and jetting (*34*, *35*). Dripping produces drops close to the entrance of the collection tube, within a single orifice diameter, analogous to a dripping faucet. Droplets produced by dripping are typically highly monodisperse. In contrast, jetting produces a long jet that extends three or more orifice diameters downstream into the collection tube, where it breaks into drops. The jetting regime is typically quite irregular, resulting in polydisperse droplets whose radii are much greater than that of the jet. Jet formation is caused by the viscous stress of the outer fluid, whose viscosity, η_{OF}, is typically 10 times greater than that of the inner fluids in our experiments. Thus, viscous effects dominate over inertial effects, resulting in a low Reynolds number. The formation of double emulsions is similar to that of single emulsions; however, there are two fluids flowing coaxially, each of which can form drops through either mechanism.

The size distribution of the double emulsions is determined by the breakup mechanism, whereas the number of innermost droplets depends on the relative rates of drop formation of the inner and middle fluids (fig. S1). When the rates are equal, the annulus and core of the coaxial jet break simultaneously, generating a

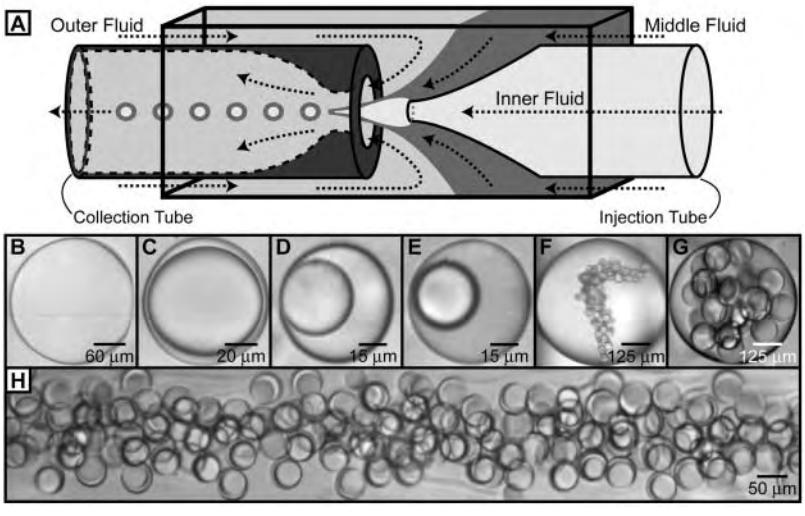


Fig. 1. Microcapillary geometry for generating double emulsions from coaxial jets. (**A**) Schematic of the coaxial microcapillary fluidic device. The geometry requires the outer fluid to be immiscible with the middle fluid and the middle fluid to be in turn immiscible with the inner fluid. The geometry of the collection tube (round tube on the left) can be a simple cylindrical tube with a constriction, as shown here, or it can be tapered into a fine point (not shown). The typical inner dimension of the square tube is 1 mm; this matches the outer diameter of the untapered regions of the collection tube and the injection tube. Typical inner diameters of the tapered end of the injection tube range from 10 to 50 μm. Typical diameters of the orifice in the collection tube vary from 50 to 500 μm. (**B** to **E**) Double emulsions containing only one internal droplet. The thickness of the coating fluid on each drop can vary from extremely thin (less than 3 μm) as in (**B**) to significantly thicker. (**F** and **G**) Double emulsions containing many internal drops with different size and number distributions. (**H**) Double emulsion drops, each containing a single internal droplet, flowing in the collection tube. The devices used to generate these double emulsions had different geometries.

double emulsion with a single internal drop. These types of double emulsions can be generated when both fluids are simultaneously dripping (Fig. 2A) or simultaneously jetting (Fig. 2B) (movies S1 to S4). The dripping and jetting mechanisms are closely related, and the transition between them is induced by varying the flow rate of the outermost fluid Q_{OF}. The dripping regime occurs for lower Q_{OF}, and increasing Q_{OF} focuses the coaxial jet more strongly, thinning the inner stream, which leads initially to smaller double emulsion drops. The radius of the outer emulsion drops, R_{drop}, decreases linearly with Q_{OF} (Fig. 3, solid circles). However, increasing Q_{OF} beyond a threshold value causes the jet to abruptly lengthen, signifying the transition to the jetting regime (*36*); this results in a discontinuous increase in R_{drop} (Fig. 3, solid triangles). In contrast, the radius of the coaxial jet, R_{jet}, measured near the exit orifice decreases monotonically through the transition (Fig. 3, half-filled symbols). As a result, the frequency of droplet production decreases.

We can construct a simple physical model for these phenomena. The classic mechanism of drop formation from a cylindrical fluid thread is through the Rayleigh-Plateau instability (*37*, *38*). This instability is driven by interfacial tension and reduces the total surface area of a fluid thread by breaking it into drops. Thus, fluid cylinders are unstable to axisymmetric perturbations with wavelengths larger than several times the radius of the cylinder itself. Although this is strictly valid only for a jet made up of a single fluid, in our experiments we matched the viscosities of the innermost and middle fluids ($\eta_{IF} = \eta_{MF}$), ensuring that the velocity profile of the coaxial jet was equivalent to that of a single fluid, and we applied the same considerations to the coaxial fluid cylinder. We also assumed that the fluids are Newtonian. The growth rate of a perturbation is determined by its velocity perpendicular to the interface, $v_\perp \sim \gamma/\eta_{OF}$, where γ is interfacial tension, leading to a droplet pinch-off time of $t_{pinch} = CR_{jet}\eta_{OF}/\gamma$, where C is a constant that depends on the viscosity ratio. Numerical calculations give $C \approx 20$ when $\eta_{IF}/\eta_{OF} = 0.1$ (*39*). Again, this is strictly valid for a cylinder of a single fluid; for a coaxial fluid thread, the effective interfacial tension may be modified by the fact that there are two interfaces. However, the Rayleigh-Plateau instability cannot occur until the length of the jet has grown to be comparable to its radius. This occurs at growth time $t_g \sim R_{jet}^3/Q_{sum}$, where Q_{sum} is the net flow rate of the two inner fluids. If $t_{pinch} < t_g$, droplets will be formed as soon as the jet is large enough to sustain the instability, which will occur right at the outlet; this leads to the dripping regime (Fig. 2A). If $t_{pinch} > t_g$, the jet will grow faster than the droplets can form; this leads to the jetting regime, where the droplets are formed downstream (Fig. 2B). We thus define an effective capillary number of the interface

$$Ca = \frac{t_{pinch}}{t_g} = \frac{\eta_{OF} Q_{sum}}{\gamma R_{jet}^2} = \frac{\eta_{OF} v}{\gamma} \qquad (1)$$

where v is the downstream velocity of the inner fluids. This equation governs the transition between dripping and jetting, which occurs when $Ca \sim 1$. Expressing the control parameter in terms of a capillary number captures the physical picture that the transition from dripping to jetting occurs when viscous stresses on the interface caused by the fluid flow become so large that the Rayleigh-Plateau instability is suppressed in these tube geometries.

Based on this simple physical picture, we can determine R_{jet} and R_{drop} as a function of Q_{OF} by considering the velocity flow profiles in both regimes. However, these velocity profiles evolve as the fluids enter and move downstream through the collection tube, necessitating a different treatment for each mechanism.

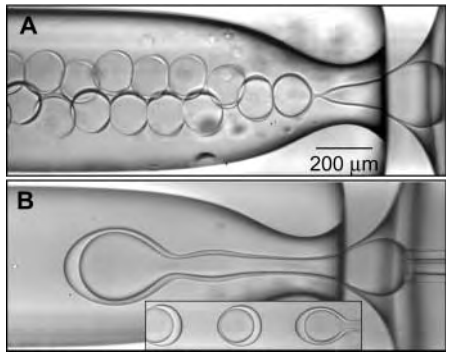


Fig. 2. Steady-state drop formation mechanisms that result in monodisperse double emulsions with a single internal droplet. (**A**) Dripping and (**B**) jetting. In both cases, the rate of drop formation is the same for the inner and middle fluids. The transition between dripping and jetting is controlled by Q_{OF} for fixed total flow rates of the inner and middle fluids. The double emulsions in (A) have a polydispersity less than 1% and those in (B) have a polydispersity of about 3%. (Inset at bottom) Pinch-off of the double emulsion drops from the coaxial jet. The same device and same fluids were used for the experiments described in Figs. 2 and 3. The outer fluid used was silicon oil with $\eta_{OF} = 0.48$ Pa·s, the middle fluid was a glycerol-water mixture with $\eta_{MF} = 0.05$ Pa·s, and the inner fluid was silicon oil with $\eta_{IF} = 0.05$ Pa·s. The interfacial tension for the aqueous mixture and silicon oil without surfactants is approximately 20 mN/m. The flow rates of the fluids in (A) are $Q_{OF} = 2500$ μl hour^{-1}, $Q_{MF} = 200$ μl hour^{-1}, and $Q_{IF} = 800$ μl hour^{-1}. The flow rates in (B) are $Q_{OF} = 7000$ μl hour^{-1}, $Q_{MF} = 200$ μl hour^{-1}, and $Q_{IF} = 800$ μl hour^{-1}. We applied these flow rates with stepper-motor–controlled syringe pumps (Harvard Apparatus, Holliston, MA).

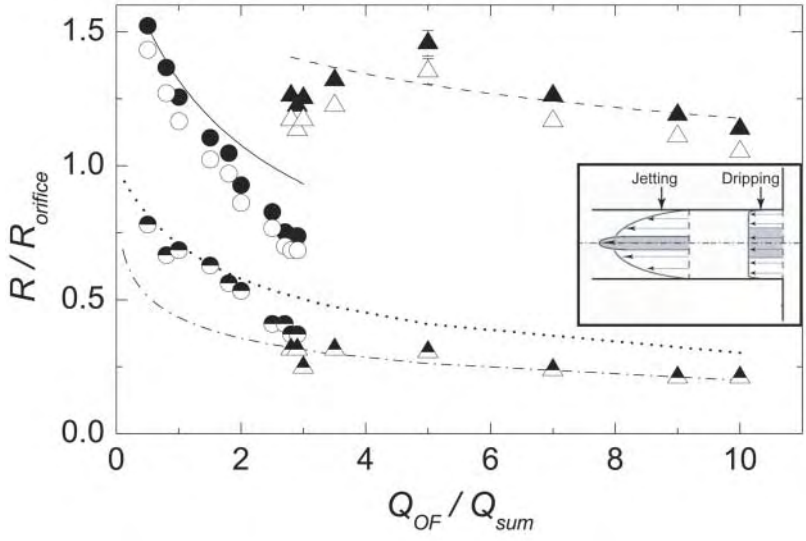


Fig. 3. Drop and jet radii versus the flow rates. The drop and jet radii are scaled by the radius of the orifice and we plotted as a function of Q_{OF}, scaled by Q_{sum}, where $Q_{sum} = Q_{MF} + Q_{IF}$. Solid circles represent the drop diameter in the dripping regime, and open circles represent the inner droplet diameter in the dripping regime. Solid triangles represent the drop diameter in the jetting regime, and open triangles represent the inner droplet diameter in the jetting regime. The solid line represents the results of the model that predicts the drop size in the dripping regime, and the dashed line represents the results of the model that predicts drop size in the jetting regime. Half-filled circles represent the jet radius in the dripping regime, and half-filled triangles represent the jet radius in the jetting regime. The dotted line is the jet radius predicted for a flat velocity profile, and the dash-dotted line is the jet radius predicted for a parabolic velocity profile. The jet radius was consistently measured at the constriction in the collection tube. (Inset) The evolution of a flat velocity profile into a parabolic velocity profile. The gray core represents the coaxial jet and the labels show the drop formation mechanism that is associated with each velocity profile. For all of the experiments, Q_{sum} was fixed at 1000 μl hour^{-1} and Q_{IF}/Q_{MF} was fixed at 4.

When the coaxial jet first enters the exit orifice, the velocity profile of the fluids is approximately flat across the channel (Fig. 3, inset). It remains this way for a distance comparable to the orifice radius times the Reynolds number. Thus, in the dripping regime, where drops form very close to the orifice, mass flux is related to cross-sectional area

$$\frac{Q_{sum}}{Q_{OF}} = \frac{\pi R_{jet}^2}{\pi R_{orifice}^2 - \pi R_{jet}^2} \quad (2)$$

where $R_{orifice}$ is the radius of the exit orifice. The values of $R_{jet}/R_{orifice}$ predicted from Eq. 2, with no adjustable parameters, are in good agreement with the measured values in the dripping regime (Fig. 3, dotted line). Comparison of the measured radii of the drops and the coaxial jet shows that $R_{drop} = 1.87\ R_{jet}$. This empirical relationship is consistent with theoretical calculations (*40*) of R_{drop} for the breakup of an infinitely long cylindrical thread in an ambient fluid for $\eta_{IF}/\eta_{OF} = 0.1$. From this we can predict R_{drop}, which is in good agreement with the data (Fig. 3, solid line). The small discrepancy near the transition probably results from deviation from a flat velocity profile.

In the jetting regime, where drops are formed well downstream, the fluid flow has evolved into the classic parabolic velocity profile of laminar pipe flow. The viscosity difference between the fluids causes the inner coaxial jet to develop a different velocity profile from that of the outer fluid. The full profiles can be determined by solving the Navier-Stokes equations in the low Reynolds number limit (SOM text II). The jet expands slightly as the collection tube widens, modifying the flow profiles; therefore, we assume a constant radius and determine the dependence of the fluid flow rates on the viscosities and $R_{jet}/R_{orifice}$ (eq. S1). We can thus predict the jet radii using no adjustable parameters, and obtain good agreement with values measured before the jet has expanded (Fig. 3, dashed-dotted line).

Drop formation in the jetting regime is generally irregular, leading to more polydisperse size distribution. However, quite stable droplet formation can be achieved, occurring at a fixed location on the jet. We believe that this condition occurs because of the geometry of our collection tube, whose diameter rapidly expands after the narrow orifice. This leads to an expansion of the jet diameter and a concomitant decrease in the velocity. As soon as *Ca* decreases sufficiently to sustain the Rayleigh-Plateau instability, droplets rupture, fixing the location and resulting in quite monodisperse droplets. In the jetting regime, the frequency of rupture decreases, producing droplets that are quite a bit larger than the size of the jet (Fig. 3). It takes time to fill these larger drops, and we can use mass conservation of the dispersed phases and the characteristic time scale for drop breakoff to obtain $Q_{sum} = (4/3)\pi R_{drop}^3/t_{pinch}$. Solving for drop radius gives,

$$R_{drop} = \left(\frac{15 Q_{sum}}{\pi} \frac{R_{jet}\eta_{OF}}{\gamma}\right)^{1/3} \quad (3)$$

This prediction yields excellent agreement with our results, as shown by the dashed line in Fig. 3. Again, there is some discrepancy between the model and data in the region of the transition, which is not described by Eq. 3. However, the predictive capability of these models provides important guidance in creating double emulsions of a desired size.

The ability to produce precision double emulsions creates many opportunities to fabricate new materials. We can control all three flow rates independently, which allows us to vary the number of internal droplets and the shell thickness. In addition, the inner and outer fluids never come into contact, facilitating great flexibility in devising schemes for encapsulation. One of the major benefits of this technique is the ability to fabricate core-shell structures, and these can easily be made into capsules. As an illustrative example, we fabricated a rigid spherical shell by photopolymerizing a polymer [Norland Optical Adhesive (NOA), Norland Products, Cranbury, NJ] in the middle fluid. We diluted the adhesive by 30% with acetone to decrease its viscosity. After generating the double emulsions, we cured the shells with an ultraviolet (UV) light source for approximately 10 s as the double emulsions traversed the collection tube. Brightfield images of the double emulsions and the resulting solid shells are shown in Fig. 4, A and B, respectively. To confirm that we indeed formed solid shells, we crushed the spheres between two microscope cover slides. A scanning electron micrograph (SEM) image confirmed the cracked polymer shells (Fig. 4C).

We can also use our device to make polymer vesicles, or polymerosomes (*41*). We created double emulsions with a single internal drop and dissolved diblock copolymers in the intermediate fluid. We used a volatile fluid as the intermediate phase and subsequently evaporated it, thereby forming polymerosomes. This method is a modified solvent evaporation technique (*15*) and is similar to a technique that has been used to produce solid polymer spheres from double emulsions (*14–16*, *19–21*). Because we can precisely control the double emulsion drop morphology, we can produce highly controllable polymerosomes. To illustrate this, we generated water-in-oil-in-water double emulsions, with a middle fluid composed of 70% toluene and 30% tetrahydrofuran (THF). This middle fluid serves as the carrier for a diblock copolymer, poly(butyl acrylate)-b-poly(acrylic acid) (PBA-PAA) (Fig. 4D) (*42*). As the solvent evaporates, the amphiphilic polymers self-assemble into layers on both interfaces, forming polymerosomes. The concentrations of the diblocks were not precisely adjusted; thus, the polymerosomes were not strictly unilamellar. However, the polymer bilayers were very thin, making it difficult to resolve them in brightfield microscopy. Thus, a phase-contrast image of a typical polymer vesicle is shown in Fig. 4E. To confirm that these structures are indeed polymerosomes, we deflated them with osmotic stress by adding a sucrose solution (0.1 M) to the continuous fluid. The polymerosome from Fig. 4E deflated in this way is shown in Fig. 4F. This

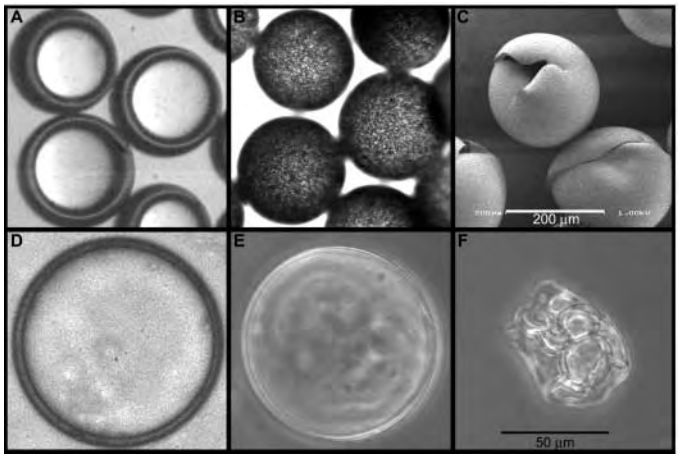


Fig. 4. Core-shell structures fabricated from double emulsions generated in our device. (**A**) Optical photomicrograph of the water-in-oil-in-water double emulsion precursor to solid spheres. The oil consists of 70% NOA and 30% acetone, with a viscosity of approximately 50 mPa·s. (**B**) Optical photomicrograph of rigid shells made by cross-linking the NOA by exposure to UV light. (**C**) SEM of the shells shown in (B) after they have been mechanically crushed. The scale bar in (C) also applies to (A) and (B). (**D**) Brightfield photomicrograph of the water-in-oil-in-water double emulsion precursor to a polymer vesicle. The oil phase consists of a mixture of toluene and tetrahydrofuran at 70/30 v/v with dissolved diblock copolymer (PBA-PAA) at 2% w/v. The viscosity of this mixture was approximately 1 mPa·s. (**E**) Phase-contrast image of the diblock copolymer vesicle after evaporation of the organic solvents. (**F**) Phase-contrast image of the deflated vesicle after osmotic stress was applied through the addition of 0.1 M sucrose to the outer fluid. The scale bar in (F) also applies to (D) and (E).

technique has a sizeable advantage: The inner and outer fluids remain totally separate, providing for efficient and robust encapsulation. In addition to polymerosomes, it should also be possible to form liposomes from phospholipids in the same manner. Alternatively, other methods to produce robust encapsulants include surface-induced polymerization of either the inner or outer interface or temperature-induced gelation of the inner or middle fluid.

Our microcapillary fluidic device is truly three dimensional, completely shielding the inner fluid from the outer one. It can generate double emulsions dispersed in either hydrophilic or hydrophobic fluids. Its production of double emulsion droplets is limited by the drop formation frequency, which varies between approximately 100 and 5000 Hz. Increasing the production rate requires the operation of parallel devices; for this, a stamping technique (*43*) would be highly desirable. However, an operational device would require control of the wetting properties of the inner channels. Alternatively, a hybrid device incorporating capillary tubes into the superstructure may offer an alternate means of making devices to produce double emulsions.

References and Notes

1. K. J. Lissant, *Emulsion and Emulsion Technology* (Marcel Dekker, New York, 1974), vol. 6.
2. C. Laugel, A. Baillet, M. P. Y. Piemi, J. P. Marty, D. Ferrier, *Int. J. Pharm.* **160**, 109 (1998).
3. C. Laugel, P. Rafidison, G. Potard, L. Aguadisch, A. Baillet, *J. Controlled Release* **63**, 7 (2000).
4. M. Gallarate, M. E. Carlotti, M. Trotta, S. Bovo, *Int. J. Pharm.* **188**, 233 (1999).
5. B. Raghuraman, N. Tirmizi, J. Wiencek, *Environ. Sci. Technol.* **28**, 1090 (1994).
6. A. Edris, B. Bergnstahl, *Nahrung/Food* **45**, 133 (2001).
7. A. Benichou, A. Aserin, N. Garti, *Polym. Adv. Technol.* **13**, 1019 (2002).
8. N. Vasishtha, H. W. Schlameus, *Microencapsulation of Food Ingredients*, P. Vilstrup, Ed. (Leatherhead Food International, Leatherhead, UK, 2001).
9. D. Lee *et al.*, *J. Dispersion Sci. Technol.* **23**, 491 (2002).
10. M. Lee, S. Oh, S. Moon, S. Bae, *J. Colloid Interface Sci.* **240**, 83 (2001).
11. V. Muguet *et al.*, *J. Controlled Release* **70**, 37 (2001).
12. S. Omi, K. Katami, T. Taguchi, K. Kaneko, M. Iso, *J. Appl. Polym. Sci.* **57**, 1013 (2003).
13. L-Y. Chu *et al.*, *J. Colloid Interface Sci.* **265**, 187 (2003).
14. R. Alex, R. Bodmeier, *J. Microencapsul.* **7**, 347 (1990).
15. S. Cohen, T. Yoshioka, M. Lucarelli, L. H. Hwang, R. Langer, *Pharm. Res.* **8**, 713 (1991).
16. N. Nihant, C. Schugens, C. Grandfils, R. Jérôme, P. Teyssié, *Pharm. Res.* **11**, 1479 (1994).
17. R. Cortesi, E. Esposito, G. Luca, C. Nastruzzi, *Biomaterials* **23**, 2283 (2002).
18. A. Lamprecht, H. Yamamoto, H. Takeuchi, Y. Kawashima, *J. Controlled Release* **98**, 1 (2004).
19. H. K. Kim, T. G. Park, *J. Controlled Release* **98**, 115 (2004).
20. M. Chiba, J. Hanes, R. Langer, *Biomaterials* **18**, 893 (1997).
21. Y. Yamaguchi *et al.*, *J. Controlled Release* **81**, 235 (2002).
22. C. Goubault *et al.*, *Langmuir* **17**, 5184 (2001).
23. S. Okushima, T. Nisisako, T. Torii, T. Higuchi, *Langmuir* **20**, 9905 (2004).
24. I. G. Loscertales *et al.*, *Science* **295**, 1695 (2002).
25. M. F. Ficheux, L. Bonakdar, F. Leal-Calderon, J. Bibette, *Langmuir* **14**, 2702 (1998).
26. R. Mezzenga, B. M. Folmer, E. Hughes, *Langmuir* **20**, 3574 (2004).
27. J. Jiao, D. G. Rhodes, D. J. Burgess, *J. Colloid Interface Sci.* **250**, 444 (2002).
28. K. Pays, J. Giermanska-Kahn, B. Pouligny, J. Bibette, F. Leal-Calderon, *Phys. Rev. Lett.* **87**, 178304 (2001).
29. L. Wen, K. D. Papadopoulos, *J. Colloid Interface Sci.* **235**, 398 (2000).
30. C. H. Villa, L. B. Lawson, Y. Li, K. D. Papadopoulos, *Langmuir* **19**, 244 (2003).
31. H. González-Ochoa, L. Ibarra-Bracamontes, J. L. Arauz-Lara, *Langmuir* **19**, 7837 (2003).
32. S. L. Anna, N. Bontoux, H. A. Stone, *Appl. Phys. Lett.* **82**, 364 (2003).
33. I. Cohen, H. Li, J. L. Hougland, M. Mrksich, S. R. Nagel, *Science* **292**, 265 (2001).
34. B. Ambravaneswaran, H. J. Subramani, S. D. Phillips, O. A. Basaran, *Phys. Rev. Lett.* **93**, 034501 (2004).
35. C. Clanet, J. C. Lasheras, *J. Fluid Mech.* **383**, 307 (1999).
36. The exact value of the flow rate at which this transition occurs will depend on the details of the tube geometry.
37. J. Plateau, *Acad. Sci. Bruxelles Mem.* **23**, 5 (1849).
38. L. Rayleigh, *Proc. R. Soc. London* **29**, 71 (1879).
39. T. R. Powers, D. Zhang, R. E. Goldstein, H. A. Stone, *Phys. Fluids* **10**, 1052 (1998).
40. S. Tomotika, *Proc. R. Soc. London Ser. A* **150**, 322 (1935).
41. B. M. Discher *et al.*, *Science* **284**, 1143 (1999).
42. PBA-PAA has a molecular weight of 5500 g/mol with the hydrophobic block (PBA) making up 70% of the total molecular weight.
43. G. M. Whitesides, A. D. Stroock, *Phys. Today* **54**, 42 (2001).

44. We thank M. P. Brenner, G. Cristobal, M. Joanicot, M. Abkarian, I. Cohen, and S. Paruchuri for helpful discussions and comments. This work was supported by NSF (grant DMR-0243715), the Harvard Materials Research Science and Engineering Center (grant DMR-0213805), Rhodia, and Unilever.

Supporting Online Material

www.sciencemag.org/cgi/content/full/308/5721/537/DC1
SOM Text
Fig. S1
Movies S1 to S4

27 December 2004; accepted 8 March 2005
10.1126/science.1109164

Retreating Glacier Fronts on the Antarctic Peninsula over the Past Half-Century

A. J. Cook,[1]* A. J. Fox,[1] D. G. Vaughan,[1] J. G. Ferrigno[2]

The continued retreat of ice shelves on the Antarctic Peninsula has been widely attributed to recent atmospheric warming, but there is little published work describing changes in glacier margin positions. We present trends in 244 marine glacier fronts on the peninsula and associated islands over the past 61 years. Of these glaciers, 87% have retreated and a clear boundary between mean advance and retreat has migrated progressively southward. The pattern is broadly compatible with retreat driven by atmospheric warming, but the rapidity of the migration suggests that this may not be the sole driver of glacier retreat in this region.

Recent changes in the Antarctic ice sheet have been caused by many different drivers, including Holocene climatic change (*1*), increasing precipitation (*2*), and changing ocean temperatures (*3*). Recent regional climate change has also begun to have a direct and immediate effect on marginal parts of the ice sheet. One notable area is the Antarctic Peninsula, where climate has warmed by ~2°C since the 1950s (*4*). Retreat of ice shelves is already well underway (*5*) with retreat of 10 ice shelves during the latter part of the 20th century (*6*–*13*). Antarctic Peninsula glaciers drain a large volume of ice, although this flux is largely balanced by snow accumulation. However, the loss of ice shelves has caused acceleration of the glaciers that fed them (*14*, *15*), creating locally high imbalances. Removal of other areas of floating ice could further increase this imbalance and thus make a substantial contribution to sea level rise.

As part of a wider project funded by the U.S. Geological Survey that will record coastal change for the whole continent (*16*, *17*), we compiled maps describing changes in the ice sheet margin around the Antarctic Peninsula. The detailed cartography, including changes in glaciers and ice shelves, will be published as maps (*18*, *19*) and as digital data (*20*). Here, we present an analysis of this comprehensive time-series data set describing the extent of the 244 marine glaciers draining the Antarctic Peninsula ice sheet and those on associated islands. All of these glaciers calve directly into the sea but comprise both tidewater glaciers and glacier ice shelves (*21*). We excluded composite ice shelves from the study because their behavior is already well documented.

The data sources (table S1) used to compile the maps included over 2000 aerial photographs dating from 1940 to 2001 and more than 100 satellite images from the 1960s onward (the example in Fig. 1 is of

[1]British Antarctic Survey, Natural Environment Research Council, Madingley Road, Cambridge CB3 0ET, UK. [2]U.S. Geological Survey, 926A National Center, 12201 Sunrise Valley Drive, Reston, VA 20192–0002, USA.

*To whom correspondence should be addressed. E-mail: acook@bas.ac.uk

two satellite image extracts revealing glacier retreat between 1986 and 2001). The glacier fronts at different dates were mapped onto a common reference in a Geographical Information System (Arc/Info). This common reference was a mosaic of Landsat images (*22*) mainly acquired during 1986 to 1989. Absolute positional accuracy was limited by the georeferencing precision of the mosaic (±160 m root mean square) (*22*). Comparison between dates was constrained by the 30-m resolution of the common reference and the nature of the source material, but 86% of the coastal data was determined to have a better than 120-m positional accuracy. The source material available did not allow the capture of glacier changes at regular or uniform intervals across the peninsula but resulted in glaciers with coastlines from many different time periods, many of which overlapped. In addition, many glacier fronts in the north were observed on more than 10 occasions, whereas farther south, the observations were less frequent, with only five date samples available in some areas.

As a basis for the statistical analysis, sample lines perpendicular to the glacier fronts were drawn for each glacier. The intersections between sample lines and glacier fronts became the basis for the analysis, and the distance from each intersection to a common point on the sample line gave a measure of advance or retreat. Because of the variable interval between observations, the mean changes (in meters per year) were averaged into 5-year intervals, ensuring that glaciers with many more observations did not skew the results. This analysis resulted in more than 22,000 measurements of change over 5-year intervals (*23*).

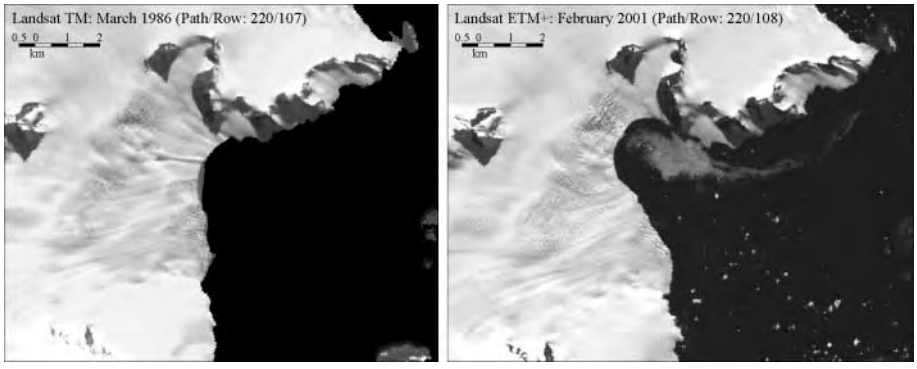


Fig. 1. Satellite image pair showing glacier retreat of Sheldon Glacier, Adelaide Island (67°32′S, 68°17′W).

The maps in Fig. 2 (and the associated statistics shown in table S2) summarize the changes in glacier fronts since the earliest records used in this study. Of the 244 glaciers, 212 (87%) have shown overall retreat since their earliest known position (which, on average, was in 1953). The other 32 glaciers have shown overall advance, but these advances are generally small in comparison with the scale of retreats observed (Table 1 shows the magnitude of changes observed). The glaciers that have advanced are not clustered in any pattern but are evenly scattered down the coast (Fig. 2).

Examination of the timing of changes along the peninsula indicates that from 1945 until 1954 there were more glaciers showing advance (62%) than those showing retreat (38%). After that time, the number in retreat has risen, with 75% in retreat in the period from 2000 to 2004. This pattern is reflected in Fig. 3, which shows that the mean rate of retreat in glaciers across the Antarctic Peninsula has been increasing since 1945. The standard deviations of retreat and advance rates over the entire domain were considerable (up to 150 m/year) but expected, due to the diverse behavior of the individual glaciers. A more useful measure of error is the likely uncertainty in the mean retreat rates, which is shown in the 95% confidence interval. The large sample size has made it possible to determine the mean behavior of glaciers on the peninsula with precision.

It has been proposed that the pattern of ice-shelf retreat on the Antarctic Peninsula is

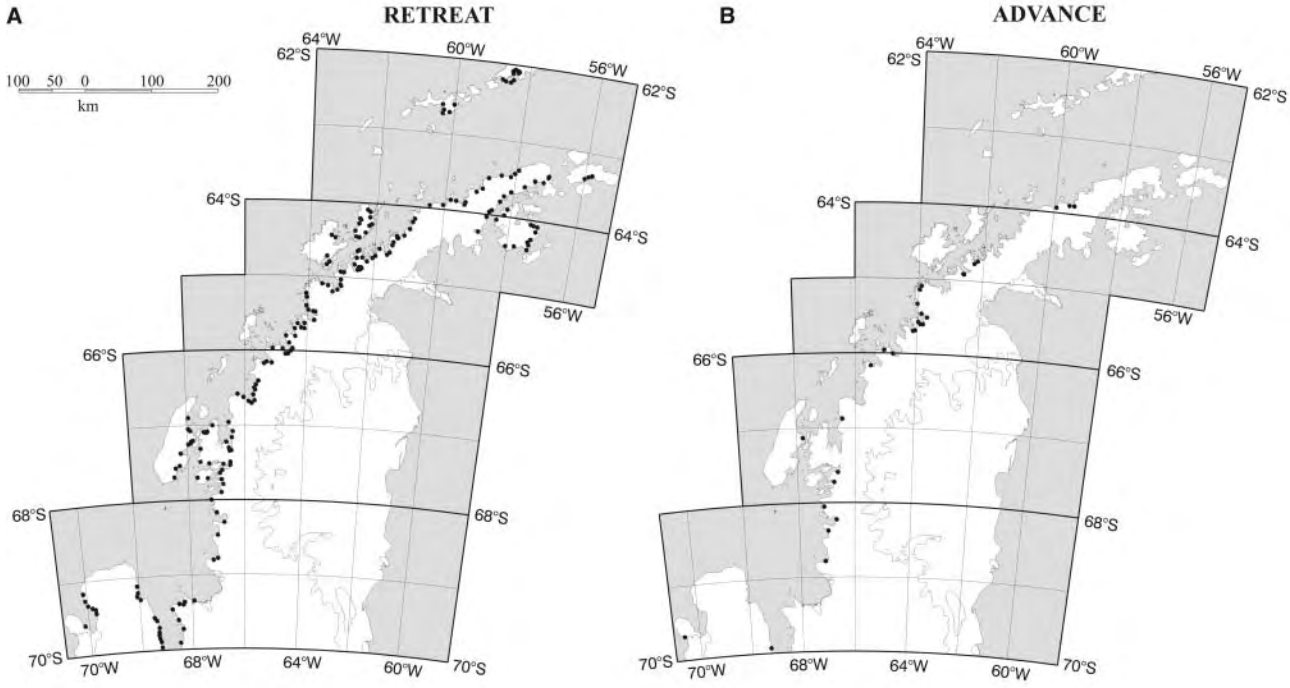


Fig. 2. Overall change observed in Antarctic Peninsula glacier fronts since earliest records.

consistent with the existence of a thermal limit on ice-shelf viability (*5*, *24*). The limit of ice shelves known to have retreated during the past 100 years is bounded by the –5° and –9°C isotherms (calculated for 2000 A.D.), suggesting that the retreat of ice shelves in this region is consistent with the observed warming trend of 3.5° ± 1.0°C per century (*24*). No ice shelves exist on the warmer side of the –9°C isotherm position, and no ice shelves on the colder side have been reported to be retreating.

Figure 4 (with accompanying statistics in table S3) reveals a similar pattern of behavior in our population of glaciers; there is a clear southerly migration of a rather abrupt transition from mean advance to mean retreat. Before 1950, glaciers north of 64°S (sector 1) were already retreating when glaciers in sectors 2, 3, and 4 were, on average, advancing. The glaciers in sector 2 transitioned from advance to retreat between 1950 and 1959, at which point only glaciers in sectors 3 and 4 were advancing. In sector 3, retreat began sometime between 1955 and at the latest, 1969 (taking into account the 95% confidence interval error bars), and in sector 4, it was between 1960 and 1969, after which glaciers in all sectors have been, on average, retreating.

In broad terms, this pattern reflects the atmospheric warming across the peninsula since the 1940s. However, a detailed examination shows that the transition from mean advance to mean retreat cannot be interpreted simply as the limit of viability for ice shelves. First, the migration of transition from mean advance to mean retreat over the period from 1955 to 1965 corresponds to a migration from the –3°C mean annual isotherm [plate 1b in (*24*)] to around –8°C in 1965. If this migration were caused by climate change, it would imply a warming much more rapid than even the maximum seasonal rate of warming: The austral winter warming at Faraday/Vernadsky Station is several times the mean annual rate (*25*). Secondly, whereas all ice shelves predicted by the limit-of-viability model appear to be retreating, the glaciers show a mixed trend with only 75% of the glaciers currently retreating.

Finally, although there is a clear reduction in the mean rate of retreat of glaciers in sectors 2, 3, and 4 in the interval from 1985 to 1994, we do not see any period of comparative cooling that could have caused this. (Figure 4 shows the mean annual temperatures for Faraday/Vernadsky binned in the same intervals; no other weather stations on the peninsula have complete air temperature records since 1945.) In addition, we have looked for correlations with summer temperature at Faraday/Vernadsky, expressed as positive degree days, and found no evidence for a causal relationship. We have analyzed satellite data from passive microwave records of sea-ice concentration for the period from 1979 to 2003 from Marguerite Bay (i.e., sector 3, Fig. 4), but these show no clear correlations with the rate of retreat of glaciers in sector 3. In particular, two periods of low sea-ice concentration (1989 to 1991 and 1998 to 2001) occurred at times when glaciers were retreating anomalously slowly and particularly rapidly, respectively. Thus, any link between retreat and sea-ice concentrations that may exist is not a simple one.

It is well known that glaciers with fully grounded marine termini exhibit unusually complex responses to changing mass balance (*26*–*28*), because in addition to the normal forcings they are subject to oceanographic forcing and subglacial topography. Indeed, several authors have suggested that the behavior of neighboring tidewater glaciers is not synchronous on decadal or century time scales (*29*). Our mixed population of tidewater and floating termini glaciers has largely shown synchronous behavior, but while we might have expected a response similar to that of ice shelves, there is no clear limit of viability to properly predict the rate of migration from mean advance to mean retreat. These glaciers must have a more complex response to climate change, perhaps to variables such as ocean temperature for which there are little data and for which we cannot test at present.

Our results indicate the existence of a clear transition between mean advance and mean retreat; a southerly migration of that transition at a time of ice-shelf retreat and progressive atmospheric warming; and a clear regime of retreat, which now exists across the Antarctic Peninsula. Future analysis of changes in all boundary conditions may reveal why the glaciers have responded in this way. Before the present retreat, the glaciers showed persistent mean advance, suggesting that the present changes could be part of longer cyclic behaviors. Such cyclic behavior, at least over decadal time scales, has been specifically ruled out for ice shelves because of their long regrowth period and evidence to suggest persistence on millennial time scales (*29*, *30*). However, it appears that in recent times this large mixed population of floating and tidewater glaciers has responded synchronously to a climatic forcing, indicating, at least statisti-

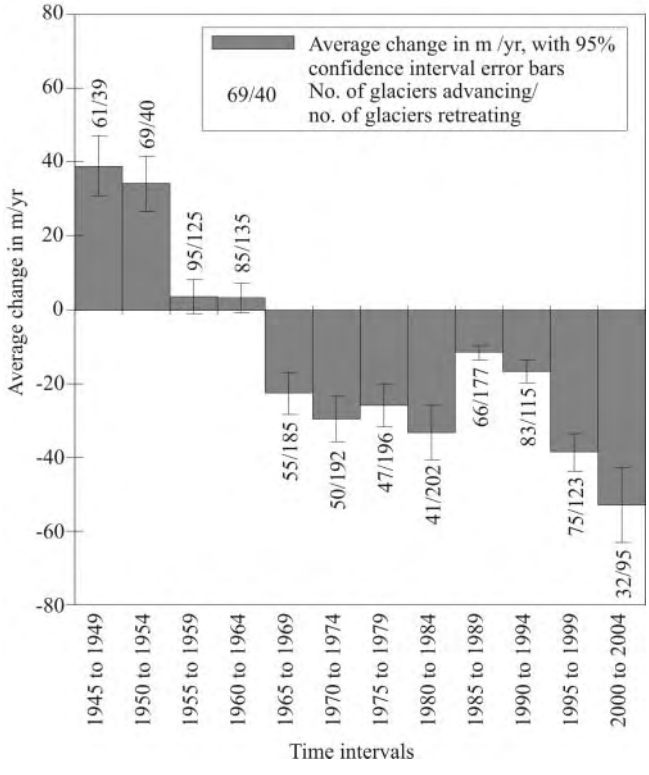


Fig. 3. Average change (with uncertainties) and average latitudes of glaciers across Antarctic Peninsula. Mean advances of 30 to 40 m/year in the 1940s and 1950s became close to stability in the 1950s and 1960s followed by an increase in rate of retreat up to the present, when retreats of 50 m/year are occurring.

Table 1. Number of glaciers showing overall advance/retreat (since earliest record) of different magnitudes. Only Sjogren Glacier at 64°14′S, 58°54′W has shown retreat of over 5 km, although this is a unique case: Until 1993, the glacier flowed into the Prince Gustav Ice Shelf, and then the ice shelf broke up and the floating glacier subsequently retreated more than 13 km.

	Advance	Retreat
Less than 500 m	27	146
500–1000 m	2	35
1000–2000 m	3	17
2000–5000 m		13
More than 5000 m		1

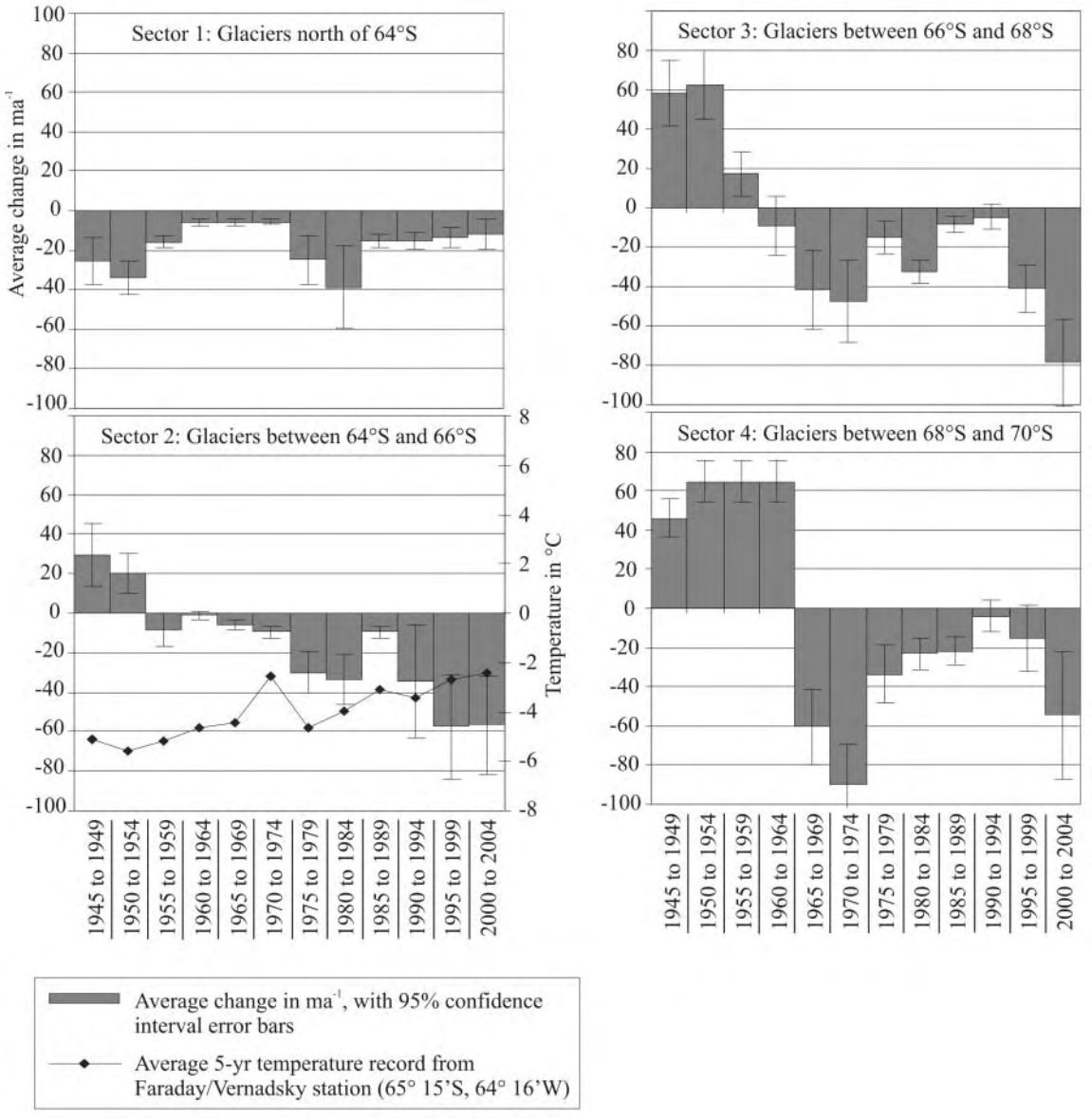


Fig. 4. Change in Antarctic Peninsula glaciers over time and by latitude. Before 1945, the limit of glacier retreat was north of 64°S; in 1955 it was in the interval between 64° and 66°S; in 1960 between 66° and 68°S; and in 1965 between 68° and 70°S. Also of note is a 10-year period in which the rate of retreat is anomalously high; it appears in all the records but is not simultaneous. In sectors 1 and 2 it appears between 1975 and 1984 and in sectors 3 and 4, between 1965 and 1974, so it appears to occur later to the north. Similarly, there is a period of noticeable reduction in the rate of retreat that occurs some time after the high retreat rates but again does not occur simultaneously between the sectors. Sector 1 shows no obvious minimum; sector 2 shows a low retreat rate between 1985 and 1989; sector 3 between 1990 and 1994; and sector 4 between 1990 and 1994, apparently occurring later to the south.

cally, that such glaciers should be predictable on decadal time scales. Because we know that the loss of floating ice shelves can cause acceleration of inland glaciers, these observations suggest that the cumulative loss of ice at the fronts of these glaciers may be leading to an increased drainage of the Antarctic Peninsula that is more widespread than previously thought.

References and Notes

1. J. O. Stone *et al.*, *Science* **299**, 99 (2003).
2. E. Mosley-Thompson, J. F. Pashkievitch, A. J. Gow, L. G. Thompson, *J. Geophys. Res.* **104**, 3877 (1999).
3. A. Shepherd, D. Wingham, T. Payne, P. Skvarca, *Science* **302**, 856 (2003).
4. D. G. Vaughan, G. J. Marshall, W. M. Connolley, J. C. King, R. Mulvaney, *Science* **293**, 1777 (2001).
5. D. G. Vaughan, C. S. M. Doake, *Nature* **379**, 328 (1996).
6. C. G. Ward, *Antarct. Sci.* **7**, 197 (1995).
7. H. Rott, P. Skvarca, T. Nagler, *Science* **271**, 788 (1996).
8. B. K. Luchitta, C. E. Rosanova, *Ann. Glaciol.* **27**, 41 (1998).
9. P. Skvarca, W. Rack, H. Rott, T. Ibarzabal y Donangelo, *Ann. Glaciol.* **27**, 628 (1998).
10. C. S. M. Doake, D. G. Vaughan, *Nature* **350**, 328 (1991).
11. A. P. R. Cooper, *Polar Rec.* **33**, 285 (1997).
12. A. J. Fox, D. G. Vaughan, in preparation.
13. T. Scambos, C. Huble, M. Fahnestock, J. Bohlander, *J. Glaciol.* **46**, 516 (2000).
14. H. Rott, W. Rack, P. Skvarca, H. de Angelis, *Ann. Glaciol.* **34**, 277 (2002).
15. H. de Angelis, P. Skvarca, *Science* **299**, 1560 (2003).
16. R. S. Williams Jr., J. G. Ferrigno, C. Swithinbank, B. K. Lucchitta, B. A. Seekins, *Ann. Glaciol.* **21**, 284 (1995).
17. R. S. Williams Jr., J. G. Ferrigno, *U.S. Geol. Surv. Fact Sheet 50–98* (1998).
18. J. G. Ferrigno, R. S. Williams Jr., J. W. Thomson, *Joint U.S. Geol. Surv.–British Antarctic Surv. Fact Sheet FS-017-02* (2002).
19. J. G. Ferrigno *et al.*, *Coastal-Change and Glaciological Maps of the Trinity Peninsula Area, Antarctica: 1843–2001* (USGS I-2600-A), in press.
20. Data are available through Antarctic Digital Database: www.add.scar.org.
21. Tidewater glaciers are defined as glaciers with grounded marine termini; glacier ice shelves are defined as the floating margins of outlet glaciers.
22. J. Sievers, A. Grindel, W. Meier, *Polarforschung* **59**(1/2), 25 (1989).
23. Although the earliest source was in 1940, only 13 glaciers (clustered in a narrow latitude band) were covered before 1945, so coastlines from these years were eliminated from the binned analysis. The latest data used for mapping coastal changes were from 2001.
24. E. M. Morris, D. G. Vaughan, *Antarct. Res. Ser.* **79**, 61 (2003).
25. D. G. Vaughan *et al.*, *Clim. Change* **60**, 243 (2003).
26. C. J. Van der Veen, *Prog. Phys. Geogr.* **26**, 96 (2002).
27. A. Vieli, J. Jania, L. Kolondra, *J. Glaciol.* **48**, 592 (2002).
28. W. T. Pfeffer, *Nature* **426**, 602 (2003).
29. E. Domack *et al.*, *Eos* **82**, 13 (2001).
30. C. J. Pudsey, J. Evans, *Geology* **29**, 787 (2001).

31. We thank R.S. Williams Jr., J.W. Thomson, K.M. Foley, C. Swithinbank, and C. Hallam for their valuable input and collaboration throughout the Coastal Change and Glaciological Maps of the Antarctic Peninsula project. We thank the U.S. Geological Survey for providing the funding to carry out this research.

Supporting Online Material

www.sciencemag.org/cgi/content/full/308/5721/541/DC1
Tables S1 to S3

18 August 2004; accepted 9 March 2005
10.1126/science.1104235

Warming of the Eurasian Landmass Is Making the Arabian Sea More Productive

Joaquim I. Goes,[1]* Prasad G. Thoppil,[2]† Helga do R Gomes,[1] John T. Fasullo[3]

The recent trend of declining winter and spring snow cover over Eurasia is causing a land-ocean thermal gradient that is particularly favorable to stronger southwest (summer) monsoon winds. Since 1997, sea surface winds have been strengthening over the western Arabian Sea. This escalation in the intensity of summer monsoon winds, accompanied by enhanced upwelling and an increase of more than 350% in average summertime phytoplankton biomass along the coast and over 300% offshore, raises the possibility that the current warming trend of the Eurasian landmass is making the Arabian Sea more productive.

The Arabian Sea's seasonally reversing monsoons drive one of the most energetic current systems in the world and the greatest seasonal variability observed in any ocean basin (*1*, *2*). It is the only ocean basin that fully reverses its circulation on a semiannual basis (*3*, *4*), a phenomenon in which the Indian Ocean, the Eurasian continent, and the Pacific Ocean play important roles (*5*). In summer (June-September), the heating of the Eurasian landmass results in low pressure over Asia, while high pressure prevails over the Indian Ocean. The geostrophically balanced airflow results in a strong topographically steered southwesterly wind and the formation of a low-level atmospheric feature called the Findlater Jet (*6*), which induces a northeastward flow of the surface current, causing strong coastal upwelling near the coasts of Somalia, Yemen, and Oman (*7*). In contrast, during the northeast monsoon (winter, November-February), the cooling of the Northern Hemispheric landmass results in high pressure over land and low pressure over the Indian Ocean, which causes a reversal in the direction of the winds from southwesterly to northeasterly (*7*). Because the reversal of the monsoons has a major influence on mixed-layer dynamics (*8*) and on physical oceanographic processes that facilitate the input of nutrients to the normally nutrient-impoverished waters of the Arabian Sea (*9*, *10*), its importance for phytoplankton growth and biogeochemical processes is profound (*11*).

From 1994 to 1996, the multinational Joint Global Ocean Flux Study (JGOFS) expeditions to the Arabian Sea helped unravel several linkages between physical forcing and carbon cycling in the northern Arabian Sea, but these were mostly on seasonal and shorter time scales (*7*, *11*, *12*). Here we present results of rapid and profound interannual changes being experienced by the Arabian Sea and, furthermore, evidence that ascribes these changes to the warming trend and the declining wintertime snow cover over the Eurasian landmass.

In 1997, the tropical Indian Ocean experienced a dipole mode (IOD) event: a pattern of zonal (east-west) variability across the ocean, with anomalously low sea surface temperatures (SSTs) off Sumatra, high temperatures in the western Indian Ocean, and accompanying wind and precipitation anomalies (*13*, *14*). This was also the year of one of the strongest El Niño events in recent history (*15*). Although uncertainty exists as to whether the dipole structure was triggered remotely by the El Niño event in the tropical Pacific or generated locally (*16*), SSTs along the entire western and central parts of the Arabian Sea were warmer than normal (*17*–*19*). Our analysis of a 7-year record of satellite ocean color data (*20*) from 1997, encompassing the period of the IOD event, revealed that concentrations of chlorophyll a in the coastal region of the western Arabian Sea (47° to 55°E, 5° to 10°N) (fig. S1) were lower than normal during the summer upwelling season of 1997 (Fig. 1A). Although satellite chlorophyll data are not available for

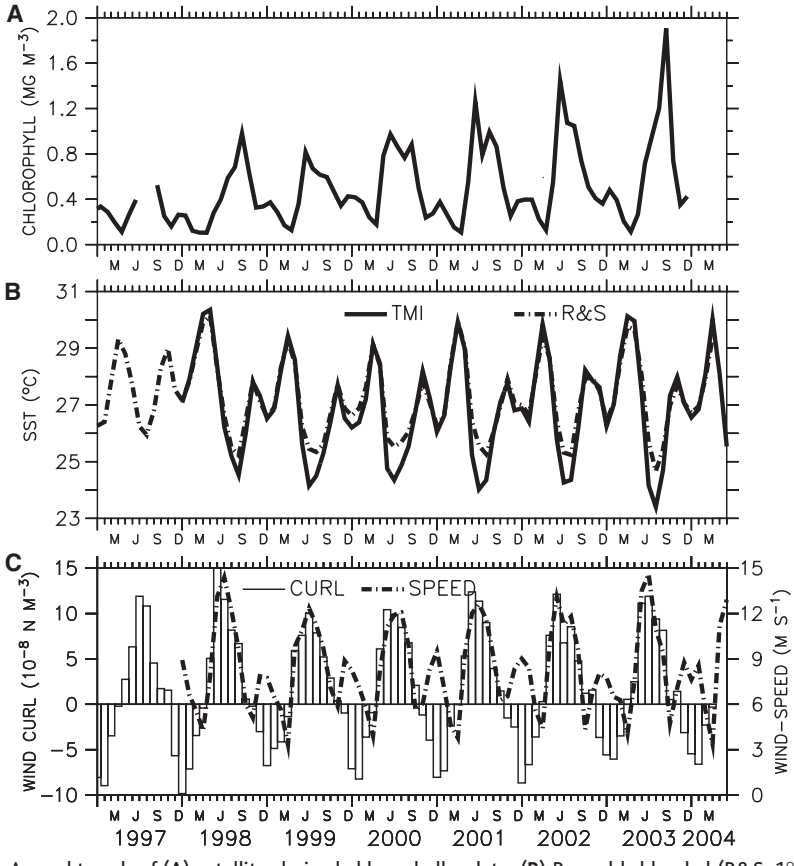


Fig. 1. Annual trends of (**A**) satellite-derived chlorophyll a data; (**B**) Reynolds blended (R&S, 1° × 1°) SSTs; and TMI (0.25° × 0.25°)–derived SSTs; and (**C**) wind stress curl values derived from NCEP-NCAR reanalysis data (open histograms) and TMI-derived wind speed for the region off the coast of Somalia (5° to 10°N and 47° to 55°E) in the western Arabian Sea. Positive wind stress curl values and lower SST values indicate upwelling, whereas negative wind stress curl values indicate downwelling. M, March; J, June; S, September; D, December.

[1]Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, ME 04575, USA. [2]Oceanography Department, Naval Postgraduate School, Monterey, CA 93943, USA. [3]Program in Atmospheric and Oceanic Sciences, University of Colorado, Boulder, CO, 80309, USA.

*To whom correspondence should be addressed. E-mail: jgoes@bigelow.org

†Present address: Department of Marine Science, University of Southern Mississippi, Stennis Space Center, MS 39529, USA.

the entire summer monsoon period of 1997 because of the premature demise of the ADEOS-1 satellite (*20*), the low chlorophyll a concentrations are explainable given that sea surface wind stress in May and June, the primary driver of upwelling during the summer monsoon in the western Arabian Sea, was much weaker than normal (*21*). The timing of the onset and the intensities of sea surface winds are both critical to the development of the Findlater Jet, which in turn is responsible for coastal divergent upwelling off the Somali coast and offshore Ekman forced upwelling off the Omani continental shelf (*3*, *4*). Coincident with the IOD event of 1997, sea surface winds (*21*) picked up only by June, almost a month later than in a normal year, followed by a peak in July that was short-lived. The impact on upwelling of the early collapse of the monsoon winds in the coastal region is clearly visible in the higher-than-normal SSTs (*22*) in June and July (Fig. 1B), an indicator of weaker upwelling that year.

In the time series record of chlorophyll a, however, the most conspicuous observation was the consistent year-by-year increase in phytoplankton biomass over the 7-year period (Fig. 1A). By the summer of 2003, chlorophyll a concentrations were >350% higher than those observed in the summer of 1997. The increase in chlorophyll a was accompanied by a year-by-year decline in summertime SSTs and cyclonic wind stress curl values (Fig. 1C) (*23*), both indicators of a progressive intensification of upwelling along the coast of Somalia resulting from a progressive strengthening of sea surface winds over the 7-year period (Fig. 1C). Upwelling off Somalia is also associated with the development of the Somali Current gyres, such as in the Great Whirl, where the vorticity balance forces an uplift of the thermocline to the left of the offshore flows (*24*, *25*).

This year-by-year increase in chlorophyll a concentrations was not confined to the coast alone but was also observed over a wider area of the western (52° to 57°E, 5°S to 10°N) Arabian Sea (Fig. 2A). Outside the region of coastal upwelling, chlorophyll a concentrations in the summer of 2003 attained values that were >300% higher than those observed in the summer of 1997. This increase in chlorophyll a was also accompanied by an intensification of sea surface winds, in particular of the zonal (east-to-west) component (Fig. 2A). It is clear from the offshore observations that the influence of southwest monsoon winds on phytoplankton in the Arabian Sea is not through their impact on coastal upwelling alone but also via the ability of zonal winds to laterally advect newly upwelled nutrient-rich waters to regions away from the upwelling zone. When colder waters are advected offshore, they cause a reduction in the latent heat flux to the atmosphere and an increase in the net heat input into the oceans. Increased heat flux into the ocean stabilizes the water column, causing the mixed layer to shoal (*26*). Thus, although sea surface winds showed a progressive year-by-year increase after 1997, mixed layer depths during the summer monsoon shallowed progressively over the 7-year period (Fig. 2B). Increased water column stability during the summer monsoon associated with a shallower mixed layer is particularly crucial for retaining phytoplankton in the euphotic layer, especially when overcast skies and insufficient light can limit phytoplankton photosynthesis and growth (*10*, *11*).

The summer monsoon winds are a coupled atmosphere-land-ocean phenomenon, whose strength is significantly correlated with tropical SSTs and Eurasian snow cover anomalies on a year-to-year basis (*27*, *28*). The intensification of the winds across the Arabian Sea during the southwest monsoon is largely governed by the land-sea thermal gradient that develops over the Arabian Sea in late spring and early summer. Therefore, the extent of winter and spring snow cover over the Eurasian landmass and the latent heat released by the sea during spring have a major impact on this land-sea thermal gradient (*29*). In general, positive snow anomalies in winter and spring can give rise to colder ground temperatures in the subsequent summer, because a substantial fraction of the available solar energy during spring and early summer goes toward melting the snow and evaporating water from the wet soil rather than toward heating the ground (*30*). Excessive snowfall in the early part of winter also tends to reduce solar radiation heating in winter by increasing the surface albedo, resulting in persistently colder temperatures over the land (*31*). Conversely, reduced snow cover over Eurasia strengthens the spring and summer land-sea thermal contrast and is considered to be responsible for the stronger southwest monsoon winds and positive rainfall anomalies over the subcontinent (*32*, *33*).

Analysis of snow cover data (*34*) for the period beginning in 1997 revealed a progressive decline of winter and spring snow cover over the Eurasian landmass (Fig. 2C), which is consistent with the mid-latitude continental warming trend reported in the Northern Hemisphere (*33*). Since 1979, the decline in snow cover has been particularly pronounced over northern Eurasia poleward of 70°N, over Western Europe, to the northeast of Russia, over southwest Asia, and over the northern Indian Himalayan Tibetan Plateau region (fig. S2). Of greatest relevance to the strength of the

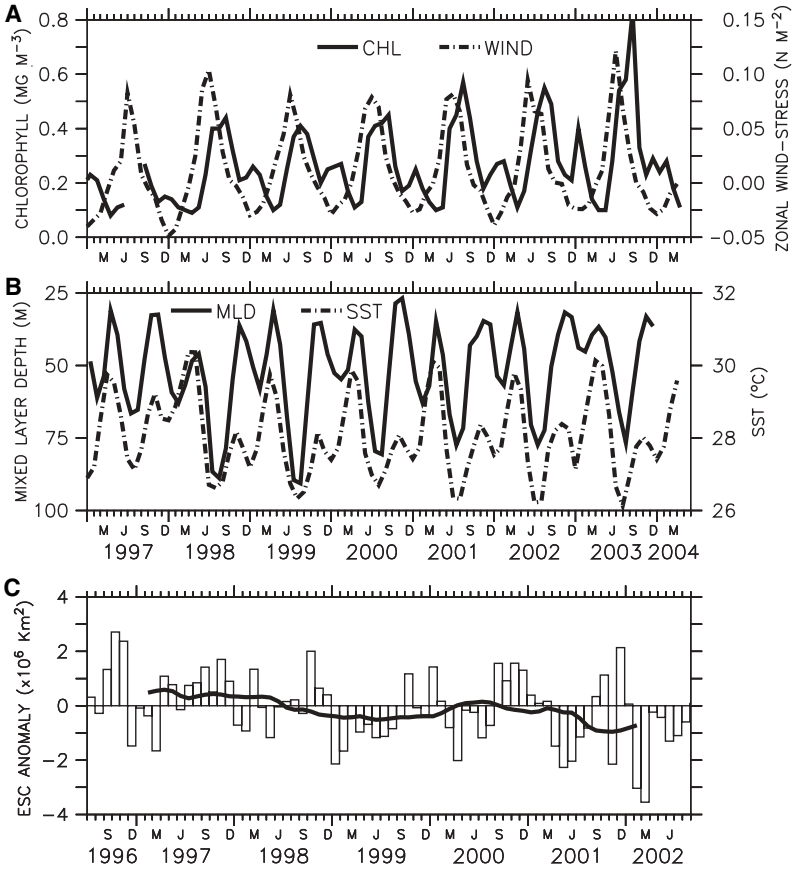


Fig. 2. Annual trends of (**A**) satellite-derived chlorophyll a data (CHL) and zonal wind stress and (**B**) mixed layer depth (MLD) and Reynolds SSTs from the region (52° to 57°E, 5°S to 10°N) in the western Arabian Sea. (**C**) Anomalies (departures from monthly means for the period between 1996 and 2002) of Eurasian snow cover (ESC). The trend line shown in bold is a 14-point moving average.

southwest monsoon winds are the latter two regions, on account of their proximity to the Arabian Sea. A plot of Eurasian snow cover extent versus wind stress data (Fig. 3A) and wind stress versus SST (Fig. 3B) for the period from late spring to midsummer (May-July) suggests that the year-by-year decline in winter and spring snow cover over Eurasia is creating conditions that are conducive to stronger winds and lower summertime SSTs across the western Arabian Sea. By regressing SST against satellite-derived chlorophyll a concentrations over this region for the summer season (Fig. 3C), we conclude that the escalating strength of sea surface winds is largely responsible for the increase in phytoplankton biomass in the western Arabian Sea over the past 7 years. The fact that disparate satellite-derived and observational data sets of SSTs and winds come together to fit into a physically consistent scenario gives us a great deal of confidence in our results.

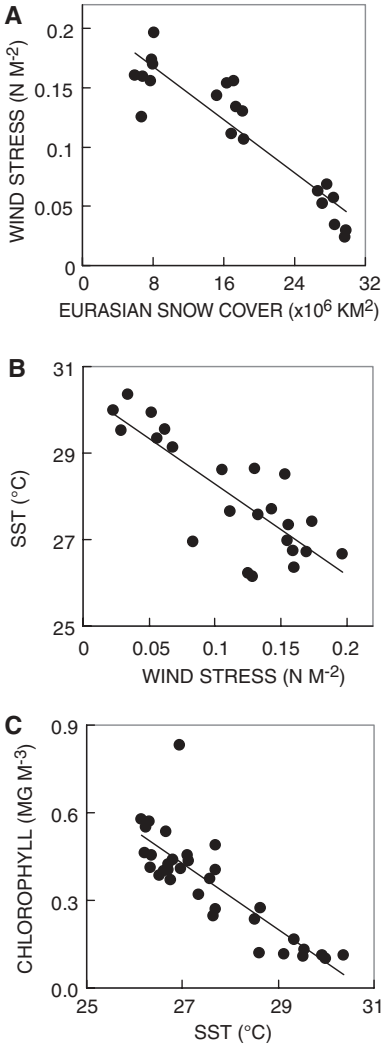


Fig. 3. Scatter plots of (**A**) resultant wind stress and Eurasian snow cover for May to July; (**B**) SSTs and resultant wind stress for May to July; and (**C**) satellite-derived chlorophyll a data and SSTs for May to September. Oceanographic data are for the region (52° to 57°E, 5°S to 10°N), and Eurasian snow cover is from the Northern Hemisphere EASE grid. Linear least-squares fits to scatter plots yielded r^2 values of 0.84, 0.66, and 0.70 ($P < 0.01$) for (A), (B), and (C), respectively.

Our findings raise the intriguing possibility that the western as well as the central regions of the Arabian Sea could witness more widespread blooms of phytoplankton if the mid-latitudinal continental warming trend and the decline in winter snow cover over the Northern Hemisphere continue. Although our findings have an immediate and important bearing on regional fisheries, the implications of a more productive Arabian Sea go far beyond that; for example, to our planet's climate. The Arabian Sea hosts a distinct, basin-wide oxygen minimum zone between 150 and 1000 m (*35*–*37*), whose presence has a substantial impact on marine elemental cycles, in particular those linked to the production of climatically relevant trace gases (*37*). The changing productivity of the Arabian Sea could thus have far-reaching consequences for the oxygen minimum zone, whose existence is regulated by a balance between the ventilation of intermediate depths and oxygen consumption during the oxidation of organic matter produced in the euphotic column (*36*, *37*).

References and Notes

1. F. A. Schott, J. P. McCreary, *Prog. Oceanogr.* **51**, 1 (2001).
2. S. Clemens, W. D. Prell, D. Murray, G. Shimmield, G. Weedon, *Nature* **353**, 720 (1991).
3. F. Schott, *Prog. Oceanogr.* **12**, 357 (1983).
4. J. C. Swallow, *Deep-Sea Res.* **31**, 639 (1984).
5. A. S. Bamzai, J. Shukla, *J. Clim.* **12**, 3117 (1999).
6. J. Findlater, *Q. J. R. Meteorol. Soc.* **95**, 362 (1969).
7. C. M. Lee, B. H. Jones, K. H. Brink, A. S. Fischer, *Deep-Sea Res. II* **47**, 1177 (2000).
8. T. G. Prasad, *J. Geophys. Res.* **109**, C03035 (2004).
9. J. D. Wiggert, R. G. Murtugudde, C. R. McClain, *Deep-Sea Res. II* **49**, 2319 (2002).
10. K. Banse, D. English, *Deep-Sea Res. II* **47**, 1623 (2000).
11. S. L. Smith, L. A. Codispoti, J. M. Morrison, R. T. Barber, *Deep-Sea Res. II* **45**, 1905 (1998).
12. M. J. R. Fasham, B. M. Balino, M. C. Bowles, *Ambio* (special report) **10**, 4 (2001).
13. N. H. Saji, B. N. Goswami, P. N. Vinayachandran, T. Yamagata, *Nature* **401**, 360 (1999).
14. P. J. Webster, A. M. Moore, J. Loschnigg, R. R. Leben, *Nature* **401**, 356 (1999).
15. J. E. Overland, N. A. Bond, J. M. Adams, *Fish. Oceanogr.* **10**, 69 (2001).
16. T. M. Shinoda, A. Alexander, H. H. Hendon, *J. Clim.* **7**, 929 (2004).
17. J. M. Slingo, H. Annamalai, *Mon. Weather Rev.* **128**, 1778 (1999).
18. R. G. Murtugudde, J. P. McCreary, A. J. Busalacchi, *J. Geophys. Res.* **105**, 3295 (2000).
19. T. G. Prasad, J. L. McClean, *J. Geophys. Res.* **109**, C02019 (2004).
20. Chlorophyll a data from November 1996 to June 1997 are reprocessed (V4.1), Level 3–binned, ADEOS-1, Ocean Color Temperature Sensor (OCTS) monthly data obtainable from the Earth Observation Center, National Space Development Agency (NASDA) of Japan. For the period from September 1997 to April 2004, we used reprocessed Sea-viewing Wide Field-of-view Sensor (SeaWiFS) (V4.1) Level 3, Global Area Coverage, monthly images from the Distributed Active Archive Center of the Goddard Space Flight Center, NASA, USA. These monthly binned products have been corrected for atmospheric light scattering and for sun angles differing from the nadir. In addition, the influence of clouds has been substantially reduced. To account for sensor degradation over time, the instrument is calibrated using internal lamps, solar diffuser observations, and lunar images, as well as vicarious methods.
21. The wind data used in this study come from two sources: the Tropical Rainfall Measurement Mission Imager (TMI) derived monthly mean (0.25° × 0.25° grid) and the monthly mean surface wind stresses (τ^x, τ^y) from the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP-NCAR) reanalysis product, calculated according to E. Kalnay *et al.* (*38*).
22. Monthly mean SSTs were obtained from two sources: the optimally interpolated (1° × 1° grid) Reynolds reanalysis product, which is a blend of Advanced High Resolution Radiometer and Comprehensive Ocean Atmosphere Data Set observations [see (*39*)]; and TMI, available from January 1998 onward. These data are processed by the Remote Sensing Systems algorithm and mapped to a 0.25° × 0.25° grid.
23. The curl of the wind stress is calculated as $\partial\tau^y/\partial x - \partial\tau^x/\partial y$, where τ^x and τ^y are the zonal and meridional wind stress, respectively.
24. G. L. Hitchcock, E. L. Ley, J. Masters, *Deep-Sea Res. II* **47**, 1605 (2000).
25. F. A. Schott, M. Dengler, R. Schoenefeldt, *Prog. Oceanogr.* **53**, 57 (2002).
26. The mixed layer depth was taken as the depth at which the temperature declined to 1°C below SST, using Expandable Bathythermograph (XBT) data optimally interpolated to a 5° × 2° longitude-latitude grid from the Joint Environmental Data Analysis Center (*40*).
27. G. A. Meehl, *Science* **266**, 263 (1994a).
28. P. J. Webster *et al.*, *J. Geophys. Res.* **103**, 14451 (1998).
29. H. F. Blanford, *Proc. R. Soc. London* **37**, 3 (1884).
30. T. P. Barnett, L. Dumenil, U. Schlese, E. Roekler, M. Latif, *J. Atmos. Sci.* **46**, 661 (1989).
31. B. Parthashastry, S. Yang, *Adv. Atmos. Sci.* **12**, 143 (1995).
32. G. A. Meehl, *J. Clim.* **7**, 1033 (1994b).
33. K. Krishna Kumar, B. Rajagopalan, M. A. Cane, *Science* **284**, 2156 (1999).
34. Snow cover extents from 1979 onward were obtained from the National Snow and Ice Data Center's CD-ROM of Northern Hemisphere EASE-Grid weekly snow cover and from ice extent data sets of R. L. Armstrong and M. J. Brodzik (available at http://nsidc.org/data/snow.html).
35. R. Sen-Gupta, S. W. A. Naqvi, *Deep-Sea Res.* **31**, 671 (1984).
36. J. M. Morrison *et al.*, *Deep-Sea Res. II* **46**, 1903 (1999).
37. S. W. A. Naqvi *et al.*, *Nature* **408**, 346 (2000).
38. E. Kalnay *et al.*, *Bull. Am. Meteorol. Soc.* **77**, 437 (1996).
39. R. W. Reynolds, T. M. Smith, *J. Clim.* **7**, 929 (1994).
40. W. B. White, S. E. Pazan, G. W. Withee, C. Noe, *Eos* **69**, 122 (1988).

41. This work is supported by grants NNG04GH50G and NNG04GM64G from NASA to J.I.G. and H.R.G. Funding from the Maine Space Grants Consortium, USA, and the Takeda Foundation, Japan, to J.I.G. and H.R.G. and a postdoctoral fellowship to T.G.P. from the Naval Postgraduate School, Monterey, USA, are gratefully acknowledged. The authors thank the Goddard Earth Sciences Data and Information Services Center/Distributed Active Archive Center, NASA, USA, and the Earth Observation Research Center, NASDA, Japan, for ocean color data from SeaWiFS and OCTS, respectively; the Climate Diagnostics Center, National Oceanic and Atmospheric Administration, USA, for the reanalysis data products; and the Remote Sensing Systems, NASA, USA, for the data from TMI. We are especially grateful to R. Armstrong and M. Brodzik of the National Snow and Ice Data Center, USA, for the Northern Hemisphere EASE-Grid snow cover data; and to K. Banse of the School of Oceanography, University of Washington, USA, for helpful comments.

Supporting Online Material

www.sciencemag.org/cgi/content/full/308/5721/545/DC1
Materials and Methods
Figs. S1 and S2

21 October 2004; accepted 18 February 2005
10.1126/science.1106610

Activation of a Phytopathogenic Bacterial Effector Protein by a Eukaryotic Cyclophilin

Gitta Coaker,[1] Arnold Falick,[2] Brian Staskawicz[1]*

Innate immunity in higher plants invokes a sophisticated surveillance system capable of recognizing bacterial effector proteins. In *Arabidopsis*, resistance to infection by strains of *Pseudomonas syringae* expressing the effector AvrRpt2 requires the plant resistance protein RPS2. AvrRpt2 was identified as a putative cysteine protease that results in the elimination of the *Arabidopsis* protein RIN4. RIN4 cleavage serves as a signal to activate RPS2-mediated resistance. AvrRpt2 is delivered into the plant cell, where it is activated by a eukaryotic factor that we identify as cyclophilin. This activation of AvrRpt2 is necessary for protease activity. Active AvrRpt2 can then directly cleave RIN4.

Plants have evolved host defense mechanisms to combat infection against a wide variety of pathogens. A gene-for-gene hypothesis was proposed to explain the plant resistance response; the hypothesis states that for a plant to be resistant to a particular pathogen, there must be matching pairs of host resistance and pathogen effector genes (*1*). Although more than 50 resistance genes have been isolated, direct binding between a resistance gene product and its corresponding pathogen effector has been demonstrated in only a few cases (*2*–*4*). Recent experimental evidence suggests an indirect mode of bacterial pathogen recognition whereby a resistance protein recognizes the biochemical alteration of a second plant protein by a bacterial effector (*5*–*8*).

Delivery of the bacterial effector AvrRpt2 by *Pseudomonas syringae* specifically induces disease resistance in *Arabidopsis* plants expressing the RPS2 resistance protein. RPS2 and RIN4 physically associate in *Arabidopsis*, and delivery of AvrRpt2 results in RIN4 elimination during pathogenesis, which supports the hypothesis that AvrRpt2 directly cleaves RIN4, thus indirectly activating RPS2 (*6*, *7*). AvrRpt2 encodes a 28-kD effector protein delivered into plant cells during infection, where it is cleaved near its amino terminus (*9*). The 21-kD carboxy-terminal product of AvrRpt2 is sufficient to trigger RIN4 elimination and RPS2 activation in planta. Amino acids 1 to 71 encode a type III secretion signal, whereas amino acids 72 to 255 encode the effector domain of AvrRpt2 (*9*). Although AvrRpt2 does not share extensive primary sequence homology to proteins of known biochemical function, the predicted secondary sequence structure of AvrRpt2$^{80–255}$ matches the catalytic core of the staphopain cysteine protease, with a putative catalytic triad composed of cysteine 122, histidine 208, and aspartate 226 (*10*). Mutations in any of these residues abolish AvrRpt2 processing, RIN4 elimination, and RPS2 recognition in *Arabidopsis* (*10*).

AvrRpt2 is not N-terminally processed in *Escherichia coli* or *P. syringae* but is processed in all eukaryotic extracts tested to date, including *Saccharomyces cerevisiae* (Fig. 1, A and B) (*9*, *11*). We hypothesize that AvrRpt2 is delivered as an inactive protease and is activated upon delivery into the plant cell (*9*, *11*). Protein extracts were prepared from *Arabidopsis* and *S. cerevisiae* and added to purified recombinant AvrRpt2 with an N-terminal hexahistidine and a C-terminal hemagglutin (HA) tag. Both crude extracts resulted in specific AvrRpt2:HA N-terminal processing (Fig. 1A). The AvrRpt2 mutant protein AvrRpt2:HA(C122A), in which alanine is substituted for the catalytic cysteine residue, was not processed (Fig. 1B). We detected two additional cleavage products after incubation with crude protein extracts (Fig. 1, A and B). These bands are likely the result of nonspecific protease activity, because they disappear upon further biochemical purification. Extensive dialysis and protein precipitation did not eliminate processing, which suggests that the factor is proteinaceous (*11*, *12*).

A biochemical approach was employed to identify the factor responsible for AvrRpt2 activation. Because *S. cerevisiae*'s genome is sequenced and contains only 6,000 genes and because there are many mutant lines available, identification of the factor in *S. cerevisiae* would be less complicated than pursuing similar experiments in *Arabidopsis*. Therefore, AvrRpt2's eukaryotic activator

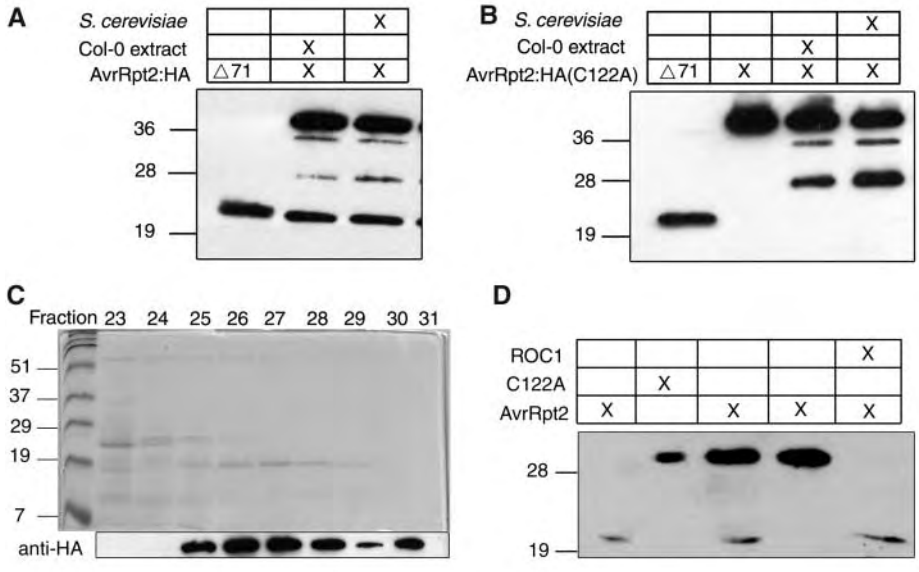


Fig. 1. Biochemical purification of AvrRpt2's eukaryotic activator. (**A** and **B**) Recombinant AvrRpt2:HA and AvrRpt2:HA(C122A) proteins were incubated with crude *S. cerevisaie* BY4730 or *Arabidopsis* Col-0 protein extract and subjected to immunoblot with monoclonal antibody to HA (anti-HA). Lane 1, nΔ71AvrRpt2:HA expressed in planta. (**C**) Biochemical purification of AvrRpt2's activator from *S. cerevisiae* BY4730 cytosol by hydrophobic interaction, anion exchange, and gel-filtration chromatography. Gel-filtration fractions surrounding activity were loaded onto a SDS-PAGE gel, stained by colloidal Coomassie blue (top panel), and subjected to anti-HA immunoblot (bottom panel) to visualize nΔ71AvrRpt2:HA. (**D**) Processing of AvrRpt2 inside *S. cerevisiae* mutants. AvrRpt2 and AvrRpt2(C122A) were expressed from the pESC vector inside *S. cerevisiae* BY4730 (lanes 1 and 2, wild type), KDY75.3b (lane 3, Δ*cpr1,6,7*, Δ*fpr1*), and KDY98.4a (lanes 4 and 5, Δ*cpr1-8*, Δ*fpr1-4*) and subjected to polyclonal anti-AvrRpt2 immunoblot. Cyclophilin mutants were complemented with *Arabidopsis* ROC1 coexpressed with AvrRpt2.

[1]Department of Plant and Microbial Biology, [2]Howard Hughes Medical Institute Mass Spectrometry Laboratory, University of California, Berkeley, CA 94720, USA.

*To whom correspondence should be addressed. E-mail: stask@nature.berkeley.edu

was purified from *S. cerevisiae*. Proteins were sequentially fractionated by hydrophobic interaction, anion exchange, and gel-filtration chromatography. Fractions were assayed for their ability to induce AvrRpt2 processing. Those possessing activity from the final purification step were analyzed by SDS–polyacrylamide gel electrophoresis (SDS-PAGE) (Fig. 1C). A single 18-kD protein correlated with activity. Mass spectrometry identified this protein as CPR1 (P14832), a single-domain cyclophilin peptidyl-prolyl cis-trans isomerase (PPIase) (fig. S1).

PPIases are chaperones and folding catalysts with the ability to catalyze the cis-trans isomerization of prolyl bonds, a rate-limiting step in protein folding (*13*). PPIases span three structurally unrelated protein families: the cyclophilins, FKPBs, and paruvulins. *S. cerevisiae* possesses eight cyclophilins (*14*). CPR1 is the yeast homolog of human cyclophilin A, localized in the cytosol and nucleus, and the primary target of the immunosuppressant drug cyclosporin A (*14*, *15*). There are seven *Arabidopsis* cytosolic single-domain cyclophilins with amino acid sequence homology ranging from 43 to 58% identity and 59 to 69% similarity to CPR1 (fig. S2). The closest *Arabidopsis* CPR1 homolog is ROC1 (At4 g38740). To verify that cyclophilin is the eukaryotic activator, AvrRpt2 processing in *S. cerevisiae* cyclophilin mutants was investigated. The *S. cerevisiae* cytosolic cyclophilin knockout did not abolish processing when AvrRpt2 was expressed in yeast, although processing was reduced in comparison with wild-type cells, which suggests that multiple cyclophilins may facilitate processing (Fig. 1D, lane 3). Yeast mutants lacking all eight cyclophilins and four FKBPs were unable to induce processing, and the inability of AvrRpt2 to process was complemented when AvrRpt2 and ROC1 were coexpressed (Fig. 1D). This verifies that the eukaryotic cofactor is cyclophilin in vivo and that ROC1 can complement CPR1 with respect to AvrRpt2 activation.

To determine whether purified CPR1 and ROC1 would enable AvrRpt2 N-terminal processing in vitro, both proteins were expressed in *E. coli* with hexahistidine N-terminal tags and purified by affinity and gel-filtration chromatography. Both CPR1 and ROC1 enabled AvrRpt2:HA processing in vitro, and mutant AvrRpt2:HA(C122A) protein was unable to process when incubated with cyclophilin (Fig. 2A). The addition of 10 μM cyclosporin A (cyclophilin inhibitor) to the in vitro reaction abolished cleavage, whereas the addition of 10 μM rapamycin (FKBP inhibitor) had no effect (Fig. 2B). The N- and C-terminal portions of AvrRpt2 were digested with trypsin and glutamyl endopeptidase, respectively, and subjected to tandem mass spectrometry. We were able to sequence up to G_{71} from the N-terminal fragment and up to G_{72} from the C-terminal fragment, verifying appropriate and specific AvrRpt2 cleavage (Fig. 2C and fig. S3). These results provide the first biochemical evidence that AvrRpt2 possesses in vitro cysteine protease activity and that AvrRpt2 activation by cyclophilin induces self-cleavage between G_{71} and G_{72}. N-terminal processing of AvrRpt2 may be necessary for protease activity or proper subcellular localization within the plant cell (*11*). Previous reports have shown that AvrRpt2 processing mutants are localized to the chloroplast rather than the membrane by monitoring AvrRpt2:GFP accumulation in protoplasts, which indicates that N-terminal processing may be required for proper localization to the membrane (*7*, *11*). However, the role of AvrRpt2's N terminus in its localization has yet to be unambiguously established.

To verify that cyclophilin and AvrRpt2 directly interact, ROC1 was expressed in *E. coli* as a glutathione S-transferase (GST) fusion protein. GST:ROC1 was incubated with AvrRpt2:HA and mutant AvrRpt2:HA(C122A) protein. Western blot analysis demonstrated that AvrRpt2:HA and AvrRpt2:HA(C122A) specifically bind to GST:ROC1 but not to GST alone (Fig. 2D). Only full-length AvrRpt2 interacts with GST:ROC1 in vitro, which suggests that this interaction is transient and that after AvrRpt2 is properly folded, the two proteins may disassociate in vitro.

Elimination of the *Arabidopsis* protein RIN4 during infection with *P. syringae* that express AvrRpt2 led to the hypothesis that RIN4 is a direct substrate for the AvrRpt2 protease (*6*, *7*). A stretch of seven amino acids surrounding AvrRpt2's cleavage site is homologous to two regions within the RIN4 protein, which suggests that AvrRpt2 may cleave RIN4 at two sites (Fig. 3A) (*16*). To confirm that RIN4 is a direct substrate of AvrRpt2 and that no other plant proteins are required for cleavage, we mixed purified recombinant AvrRpt2:HA, AvrRpt2:HA(C122A), ROC1, and RIN4 proteins in vitro. RIN4 was only cleaved in the presence of wild-type AvrRpt2:HA and ROC1, verifying that AvrRpt2 directly cleaves RIN4 after it is activated by cyclophilin (Fig. 3, B and C). The two RIN4 fragments were identified by mass spectrometric peptide mapping and tandem mass spectrometry as the N terminus of RIN4 and an internal fragment of RIN4 between sites I and II (Fig. 3C and fig. S4B). The C-terminal fragment of RIN4 was undetectable by immunoblot analysis or Coomassie staining. In addition, full-length RIN4 protein was analyzed by mass spectrometry, verifying that the C terminus was

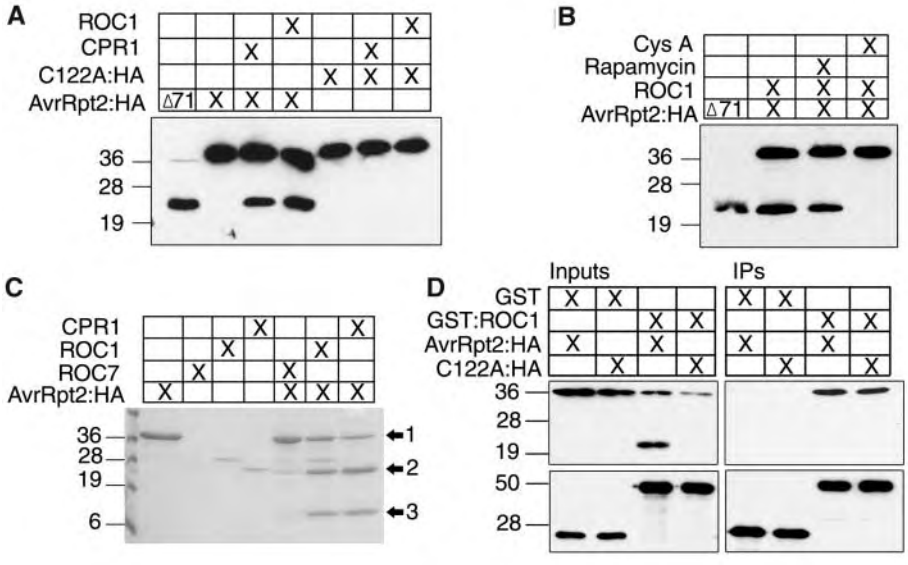


Fig. 2. Cleavage site specificity of the AvrRpt2 protease. (**A**) Recombinant *Arabidopsis* and *S. cerevisiae* cyclophilins induce AvrRpt2 processing in vitro. Recombinant AvrRpt2:HA or AvrRpt2:HA(C122A) were incubated with ROC1 and CPR1 and subjected to anti-HA immunoblot. Lane 1, nΔ71AvrRpt2:HA expressed in planta. (**B**) Cyclosporin A inhibits AvrRpt2:HA activation by ROC1. Recombinant AvrRpt2:HA was incubated with ROC1, cyclosporin A (Cys A), and rapamycin and subjected to anti-HA immunoblot. Lane 1, nΔ71AvrRpt2:HA expressed in planta. (**C**) AvrRpt2 self-processes between G_{71} and G_{72}. Recombinant AvrRpt2:HA, ROC1, ROC7, and CPR1 were incubated in various combinations and visualized by Coomassie blue. Full-length AvrRpt2(1) was cleaved in the presence of cyclophilin, generating the C terminus of AvrRpt2(2) and the N terminus of AvrRpt2(3). (**D**) AvrRpt2 and ROC1 directly interact. Recombinant AvrRpt2:HA proteins were incubated with GST:ROC1 and GST alone. GST interacting proteins were recovered by incubation with G beads. Input controls (lanes 1 to 4) and GST binding proteins (IPs) were detected by anti-HA (top panel) and anti-GST (bottom panel) immunoblots.

intact prior to cleavage by AvrRpt2 (fig. S4A). Recombinant RIN4 mutant proteins [RIN4(F9A), RIN4(F151A), and RIN4(F9A F151A)] were incubated with AvrRpt2 and ROC1, and RIN4 cleavage products were identified by mass spectrometry. RIN4(F9A) was cleaved only at site II, RIN4(F151A) was cleaved only at site I, and RIN4(F9A F151A) was not cleaved by AvrRpt2, indicating that AvrRpt2 directly cleaves RIN4 at two positions in vitro (Fig. 3B and fig. S4, C and D). We were able to identify RIN4 site I as being cleaved between G_{10} and N_{11} by tandem mass spectrometry (fig. S5). AvrRpt2 also cleaves itself at a similar position in its recognition site (fig. S3). Taken together, these data suggest that AvrRpt2 cleaves RIN4 site II between G_{152} and D_{153}.

Previous experiments indicate that the RIN4 double mutants are also not cleaved in planta after delivery of AvrRpt2 (*17*, *18*). The RIN4(F9A F151A) double mutant is not eliminated when transiently expressed with AvrRpt2 in *Nicotiana benthamiana*, whereas RIN4 single mutants are undetectable by immunoblot, which suggests that RIN4 is cleaved at two sites in planta (*17*). Furthermore, the RIN4 site I and II mutant is unable to be eliminated in *Arabidopsis* upon infection with *P. syringae* expressing AvrRpt2 (*18*). Because RIN4 negatively regulates RPS2 (*6*, *7*), RPS2-mediated resistance does not occur in *Arabidopsis* plants expressing the RIN4 double mutant upon infection with *Pseudomonas* expressing AvrRpt2 (*18*).

Together, these results suggest a novel mechanism to explain AvrRpt2's activation upon delivery into the plant cell. Spontaneous peptidyl-prolyl isomerization is a slow reaction and can constitute rate-limiting steps in protein folding. The discovery that eukaryotic cyclophilin activates AvrRpt2 not only provides a mechanism for effector activation upon delivery into the plant cell but also implicates *Arabidopsis* single-domain cyclophilins as general protein-folding catalysts.

We propose that the PPIase activity of cyclophilin activates AvrRpt2, enabling it to cleave its N terminus, localize to the membrane where RPS2 and RIN4 reside, and directly cleave RIN4 (*5*, *6*). RIN4 is a negative regulator of RPS2 (*6*, *7*). It is likely that the cleavage of RIN4 activates RPS2, because *Arabidopsis rin4* null mutants are lethal in an RPS2 background (*6*) and because the overexpression of RIN4 delays RPS2 activation (*6*, *7*). Growth of *Pseudomonas* expressing AvrRpt2 has been shown to be significantly higher than that in catalytically inactive mutants on *Arabidopsis rin4/rps2* plants (*17*). These data suggest that, in addition to RIN4, AvrRpt2 likely has additional protein targets inside the plant cell that account for its virulence activity. A set of *Arabidopsis* proteins possessing homology to AvrRpt2's processing site was identified that are cleaved by AvrRpt2 when both proteins are transiently expressed in *N. benthamiana*, and these proteins may be virulence targets of the AvrRpt2 protease (*17*). Targeting of multiple host proteins may be a common strategy employed by a number of effectors, and future studies will investigate cleavage of this set of proteins in *Arabidopsis* and their potential roles in susceptibility and resistance.

A common theme in host and pathogen proteinases is that the expression of proteolytic activity is tightly regulated, both at the level of expression and secretion and in the processing of an inactive secreted precursor to its active form (*19*, *20*). There are 6 proline residues in the prodomain and 12 residues within the effector domain of AvrRpt2 that may be important for interacting with cyclophilin. A number of bacterial effectors delivered into plant and animal cells during pathogenesis are structurally similar to cysteine proteases (*21*, *22*). Furthermore, enzyme activity of several putative bacterial effector proteases cannot be detected in vitro, which indicates that the effectors YopJ, AvrBsT, and AvrXv4 may also require host factors for activation (*22*). In light of our data, folding of bacterial effector proteases by eukaryotic protein factors may be a common mechanism during pathogenesis.

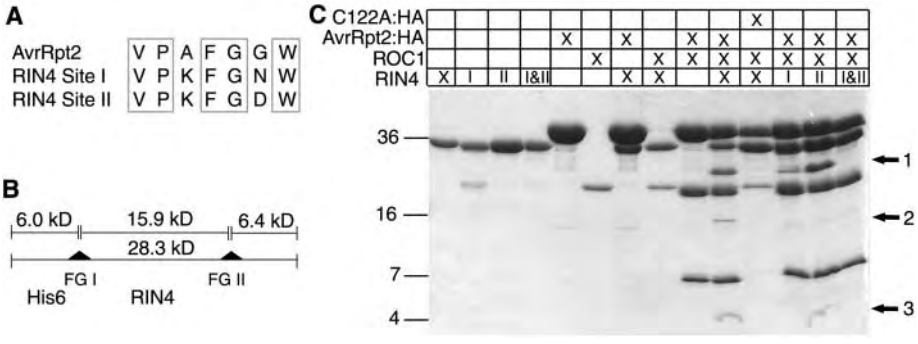


Fig. 3. RIN4 is a direct substrate for AvrRpt2. (**A**) RIN4 possesses two sites with homology to AvrRpt2's processing site. RIN4 site I is located between V_6 and W_{12}, and site II is between V_{148} and W_{154}. (**B**) Diagram of the RIN4 recombinant protein. FG I, site I; FG II, site II. (**C**) RIN4 is cleaved by activated AvrRpt2 in vitro. Recombinant AvrRpt2:HA, AvrRpt2(C122A):HA, ROC1, and RIN4 proteins were incubated in combinations in vitro, and cleavage products were visualized by Coomassie blue. I indicates RIN4(F9A), II indicates RIN4(F151A), and I&II indicates RIN4(F9A F151A) mutant proteins. RIN4 cleavage products were identified by mass spectrometry. Fragment 1 results from a single cleavage event at either site I or site II, fragment 2 is the internal portion between sites I and II, and fragment 3 is the N-terminal piece of RIN4. The C-terminal fragment of RIN4 is undetectable after AvrRpt2 cleavage at site II in vitro.

References and Notes

1. H. H. Flor, *Phytopathology* **45**, 680 (1955).
2. J. L. Dangl, J. D. Jones, *Nature* **411**, 826 (2001).
3. Y. Jia, S. A. McAdams, G. T. Bryan, H. P. Hershey, B. Valent, *EMBO J.* **19**, 4004 (2000).
4. L. Deslandes *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8024 (2003).
5. D. Mackey, B. F. Holt, A. Wiig, J. L. Dangl, *Cell* **108**, 743 (2002).
6. D. Mackey, Y. Belkhadir, J. M. Alonso, J. R. Ecker, J. L. Dangl, *Cell* **112**, 379 (2003).
7. M. J. Axtell, B. J. Staskawicz, *Cell* **112**, 369 (2003).
8. F. Shao *et al.*, *Science* **301**, 1230 (2003).
9. M. B. Mudgett, B. J. Staskawicz, *Mol. Microbiol.* **32**, 927 (1999).
10. M. J. Axtell, S. T. Chisholm, D. Dahlbeck, B. J. Staskawicz, *Mol. Microbiol.* **49**, 1537 (2003).
11. P. Jin, M. D. Wood, Y. Wu, X. Xie, F. Katagiri, *Plant Physiol.* **133**, 1072 (2003).
12. Coaker and Staskawicz, unpublished data.
13. T. Kiefhaber, R. Quaas, U. Hahn, F. X. Schmid, *Biochemistry* **29**, 3061 (1990).
14. K. Dolinski, S. Muir, M. Cardenas, J. Heitman, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13093 (1997).
15. M. Arevalo-Rodriguez, J. Heitman, *Eukaryot. Cell* **4**, 17 (2005).
16. D. A. Jones, D. Takemoto, *Curr. Opin. Immunol.* **16**, 48 (2004).
17. S. T. Chisholm *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 2087 (2005).
18. B. Day *et al.*, *Plant Cell* **17**, 1292 (2005).
19. C. Wandersman, *Mol. Microbiol.* **3**, 1825 (1989).
20. M. L. Corcoran, D. E. Kleiner, W. G. Stetler-Stevenson, *Adv. Exp. Med. Biol.* **385**, 151 (1995).
21. J. R. Alfano, A. Collmer, *Annu. Rev. Phytopathol.* **42**, 385 (2004).
22. M. B. Mudgett, *Curr. Opin. Plant Biol.* **7**, 384 (2004).

23. We thank J. Heitman for providing KDY75.3b and KDY98.4a *S. cerevisiae* strains. We thank J. Rhine and J. Thorner for many enlightening discussions. We thank all members of the Staskawicz laboratory for critical reading of this manuscript. This work was supported by Department of Energy grant DE-FG02-88ER13917 and NIH grant R01-FM069680-01, awarded to B.S. G.C. is supported by U.S. Department of Agriculture postdoctoral fellowship 019136-003.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1108633/DC1
Materials and Methods
Figs. S1 to S5
References and Notes

13 December 2004; accepted 4 February 2005
Published online 3 March 2005;
10.1126/science.1108633
Include this information when citing this paper.

ATM Activation by DNA Double-Strand Breaks Through the Mre11-Rad50-Nbs1 Complex

Ji-Hoon Lee and Tanya T. Paull*

The ataxia-telangiectasia mutated (ATM) kinase signals the presence of DNA double-strand breaks in mammalian cells by phosphorylating proteins that initiate cell-cycle arrest, apoptosis, and DNA repair. We show that the Mre11-Rad50-Nbs1 (MRN) complex acts as a double-strand break sensor for ATM and recruits ATM to broken DNA molecules. Inactive ATM dimers were activated in vitro with DNA in the presence of MRN, leading to phosphorylation of the downstream cellular targets p53 and Chk2. ATM autophosphorylation was not required for monomerization of ATM by MRN. The unwinding of DNA ends by MRN was essential for ATM stimulation, which is consistent with the central role of single-stranded DNA as an evolutionarily conserved signal for DNA damage.

In mammalian cells, DNA double-strand breaks trigger activation of the ataxia-telangiectasia mutated (ATM) protein kinase, which phosphorylates downstream targets that initiate cell-cycle arrest, DNA repair, or apoptosis. Several of these targets, including p53, Chk2, Brca1, and H2AX, function as tumor suppressors in vivo, and the phosphorylation of these factors is critical for their function after DNA damage.

The Nbs1 (nibrin) protein is also a substrate for ATM, and abrogation of Nbs1 phosphorylation inhibits checkpoint signaling during the S phase (chromosome replication) of the cell cycle (*1*). Nbs1 is part of the Mre11-Rad50-Nbs1 (MRN) complex, which is essential for DNA double-strand–break repair and genomic stability. Cells from patients with Nijmegen breakage syndrome (NBS) or ataxia telangiectasia–like disorder (ATLD) express mutant forms of the Nbs1 or Mre11 protein, respectively, and exhibit decreased levels of ATM substrate phosphorylation, particularly on Chk2 (*2*–*5*) and Smc1 (*6*, *7*), despite the presence of wild-type ATM. Thus, the MRN complex may not only be a downstream effector of ATM but also may function in activating ATM to initiate phosphorylation of cellular substrates.

MRN stimulated ATM activity in vitro toward p53, Chk2, and histone H2AX in a kinase assay with purified recombinant components (*8*). MRN and ATM associated through multiple protein-protein interactions, and MRN contributed to ATM kinase activity by increasing the affinity of ATM for its substrates. In this in vitro assay, however, there was no effect of DNA on ATM, either with or without MRN (Fig. 1A).

ATM exists in vivo as an inactive multimer that dissociates into active monomers after DNA damage or other forms of cellular stress (*9*). Thus, the ATM we studied previously may have been monomeric, either already present as a monomer in cells or converted into monomers during purification. To study the multimeric form of ATM specifically, we transfected human 293T cells with two ATM expression constructs encoding Flag- and hemagglutinin (HA)–epitope–tagged ATM, and we modified our purification procedure to preserve multimeric interactions. Sequential purification with antibodies directed against the Flag and HA epitopes yielded ATM complexes (Fig. 1B). Glycerol gradient sedimentation analysis of complexes containing both epitope-tagged forms of ATM showed that the majority of this protein fractionated as a dimer, whereas the previously purified form of ATM fractionated as a monomer (Fig. 1C).

Unlike monomeric ATM, dimeric ATM required both the MRN complex and DNA for activity. Dimeric ATM was tested for kinase activity with a glutathione S-transferase (GST) fusion protein containing residues 1 to 102 of p53, and phosphorylation was detected with a phosphospecific antibody directed against p53 protein phosphorylated on serine 15 (S15). Minimal activity was seen with ATM alone, or with ATM with MRN, whereas substantial phosphorylation was seen with ATM, MRN, and linear DNA incubated together in the kinase reaction (Fig. 1D). Similar results were seen with dimeric ATM and a GST-Chk2 substrate when an antibody that is specific for phosphothreonine 68 was used (Fig. 1E). In each case, MRN increased the amount of phosphorylated product by only two- to fivefold, whereas DNA plus MRN yielded an increase in phosphorylated product of 80- to 200-fold over ATM alone. The addition of DNA to dimeric ATM in the absence of MRN did not stimulate kinase activity (Fig. 1, D and E). Dimeric ATM also required the complete MRN complex and was not stimulated by MR complex, which lacks Nbs1 (Fig. 1F), unlike monomeric ATM, which is stimulated by both the MR and MRN (Fig. 1A). Identical results were observed with full-length Chk2 and p53 (fig. S2) (*10*).

The level of ATM stimulation by DNA was modulated by the substrate concentration (Fig. 1G and fig. S1). The increase in phosphorylation induced by DNA over that by MRN alone was greater than 60-fold with 12.5 nM GST-p53 substrate and 32-fold with 50 nM substrate, but only 18-fold with 200 nM substrate. Therefore, DNA and MRN may act on dimeric ATM by stimulating substrate recruitment, in similar fashion to how they act on monomeric ATM (*8*).

In vivo, ATM is activated by DNA double-strand breaks. To test whether DNA ends are required with dimeric ATM, we added closed circular plasmid DNA instead of linear DNA fragments. The circular DNA stimulated ATM fivefold in the presence of MRN, but when restriction enzymes were also included in the kinase reaction, the phosphorylation of p53 and Chk2 was increased by 13- to 25-fold (Fig. 2A). We did not observe any DNA sequence specificity (*11*); however, optimal stimulation of dimeric ATM required that the length of the DNA fragment be at least 1 to 2 kb (Fig. 2B).

To determine which proteins are required for DNA binding, we attached a 2.3-kb DNA fragment to magnetic beads through a biotin-streptavidin interaction. The DNA-bound beads were incubated with recombinant ATM and MRN, and we identified the bound proteins by Western blotting. MRN bound to the DNA-containing beads irrespective of the presence of ATM, whereas ATM was associated with the beads only when MRN was present (Fig. 2C). Thus, in vitro, MRN is required for the stable association of ATM with DNA. The MR complex, which lacks Nbs1, also stimulated ATM binding to DNA (Fig. 2D), but Mre11 alone did not (Fig. 2E), suggesting that the interaction between MR and ATM (*8*) is through the Rad50 component of the complex. Both MRN and MR recruited ATM to DNA (Fig. 2D), yet only MRN stimulated the kinase activity of ATM (Fig. 1F). Similarly, ATM bound equally well to 415–base pair (bp) and 2.3-kb DNA fragments in the presence of MRN (Fig. 2C), yet was only efficiently activated by the larger fragments (Fig. 2B). Thus, ATM recruitment to DNA does not appear to be sufficient for activation.

ATM autophosphorylation of Ser^{1981} (S1981) was shown to be required for the activation of ATM activity and for the damage-induced conversion of ATM dimers

Department of Molecular Genetics and Microbiology, Institute of Cellular and Molecular Biology, University of Texas at Austin, 1 University Station, A4800, Austin, TX 78712, USA.

*To whom correspondence should be addressed. E-mail: tpaull@icmb.utexas.edu

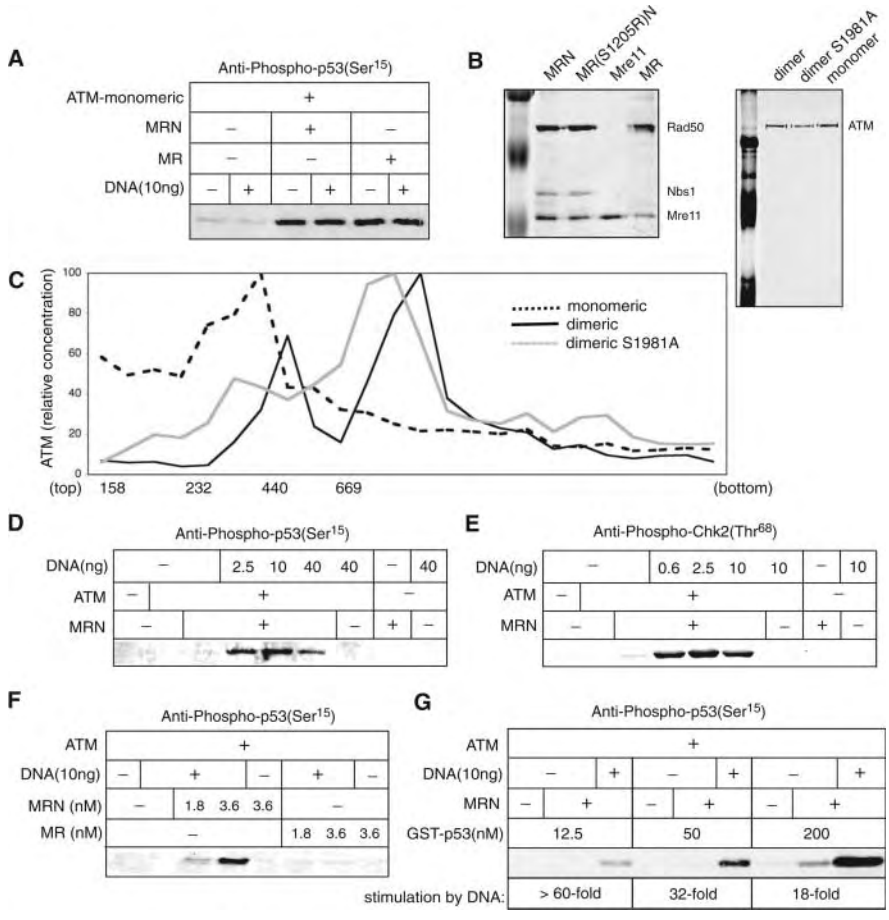


Fig. 1. Requirement of DNA and MRN for activity of dimeric ATM. (**A**) Kinase assays contained 0.125 nM monomeric ATM, 7.5 nM MRN, 7.5 nM MR, 50 nM GST-p53 (amino acids 1 to 102), and 10 ng of linear DNA. Western blots were probed with antibody to phosphoserine 15 of p53. (**B**) Coomassie-stained SDS–polyacrylamide gel electrophoresis (SDS-PAGE) of MRN, MR(S1202R)N, MR, Mre11, Flag-tagged monomeric ATM (monomer), Flag-ATM–HA-ATM dimeric complex (dimer), and S1981A dimeric complex (dimer S1981A). (**C**) Glycerol gradient sedimentation analysis of monomeric ATM, dimeric ATM, and S1981A dimeric ATM, with positions of molecular size markers as indicated. ATM concentrations are shown relative to the most concentrated fraction within each gradient. (**D**) Kinase assays with 0.2 nM dimeric ATM, 3.6 nM MRN, 50 nM GST-p53 substrate, and linear DNA, probed with antibody to phosphoserine 15 of p53. (**E**) Kinase assays with 0.2 nM dimeric ATM, 4.8 nM MRN, 200 nM GST-Chk2 substrate, and linear DNA, probed with antibody to phosphothreonine 68 of Chk2. (**F**) Kinase assays as in (D) with indicated amounts of MRN or MR (1.8 and 3.6 nM, respectively). (**G**) Kinase assays as in (D) with indicated amounts of GST-p53. Anti-GST Western blots are shown in fig. S1.

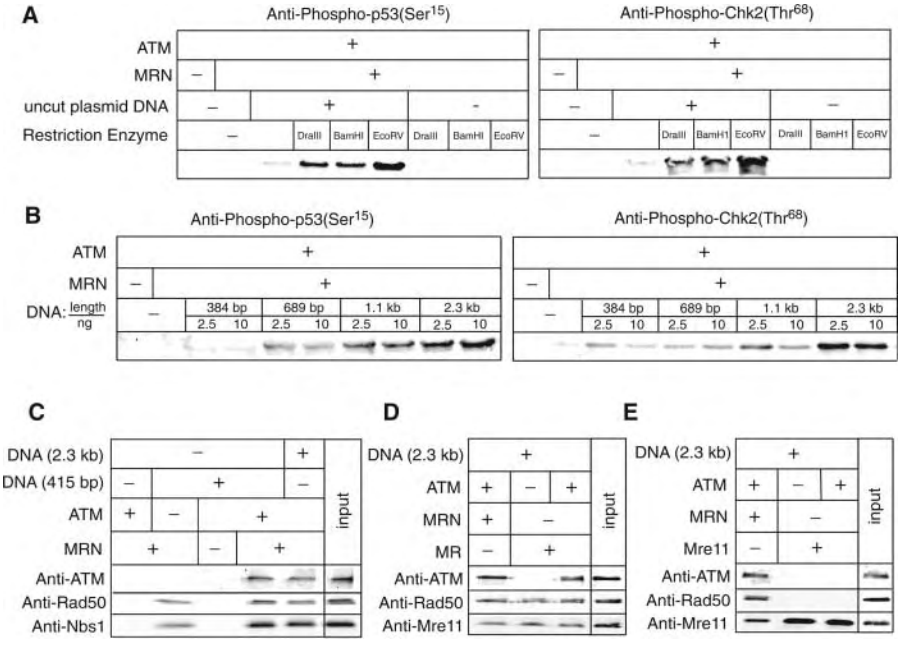


Fig. 2. Importance of DNA ends for ATM activation through MRN. (**A**) Kinase assays as in Fig. 1 are shown except with closed circular relaxed plasmid DNA (uncut plasmid DNA) instead of linear DNA fragments. One unit of Dra III, Bam HI, or Eco RV restriction enzyme was added as indicated. Each enzyme cuts a unique site in the plasmid. (**B**) Kinase assays with linear DNA of various lengths. (**C**) Biotinylated 415-bp and 2.3-kb DNA fragments were attached to streptavidin-coated magnetic beads and incubated with MRN and ATM. Associated MRN and ATM were detected by Western blotting. The input lane has one-10th of ATM and one-third of MRN used in the reaction. (**D** and **E**) Binding assays as in (C), except with MR or Mre11 compared to MRN.

to monomers (*9*). A small amount of ATM S1981 autophosphorylation in the in vitro assay was detected with a phosphospecific antibody, and it increased threefold in the presence of MRN and DNA (Fig. 3A). To test whether autophosphorylation of ATM on S1981 is necessary for the DNA stimulation of dimeric ATM in vitro, we cotransfected Flag- and HA-tagged ATM S1981A mutant (where Ser^{1981} is replaced by Ala) expression constructs and purified complexes with antibodies to Flag (anti-Flag) and anti-HA (Fig. 1B). This S1981A protein also migrated as a dimer in the glycerol gradient (Fig. 1C). The S1981A dimeric ATM, as well as the wild-type dimeric protein, responded similarly to DNA and MRN, exhibiting greater than 200-fold stimulation of both p53 and Chk2 phosphorylation by DNA and MRN (Fig. 3, A and B). Thus, in vitro, S1981 autophosphorylation is not essential for MRN-dependent stimulation of dimeric ATM by DNA.

The conversion of ATM dimers into monomers occurs in human cells after DNA damage and correlates with ATM activation (*9*). To assay the dimer-to-monomer transition in vitro, we bound the Flag-ATM–HA-ATM dimer preparation to anti-Flag conjugated to agarose beads. After washing the beads, we

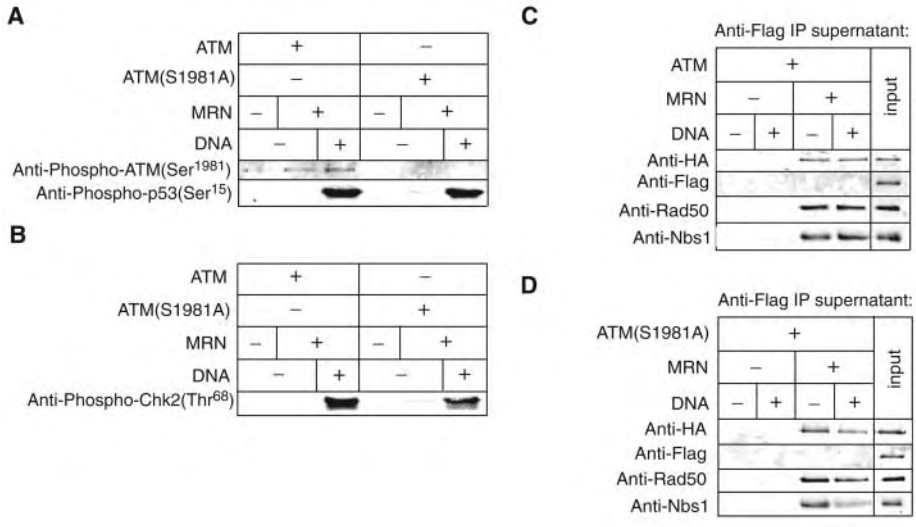


Fig. 3. ATM autophosphorylation of S1981. (**A** and **B**) Kinase assays as in Fig. 1D, with S1981A dimeric ATM compared with wild-type dimeric ATM, with 3.6 nM MRN and linear DNA as indicated. (**C**) The Flag-ATM–HA-ATM dimer was bound to anti-Flag conjugated to agarose beads. Dissociation of ATM was monitored by Western blotting of the supernatant with anti-HA and antibodies to Flag, Rad50, and Nbs1. The input lane has one-third of the ATM and one-10th of the MRN used in the reaction. (**D**) Dissociation of the S1981A dimeric ATM complex as in (C).

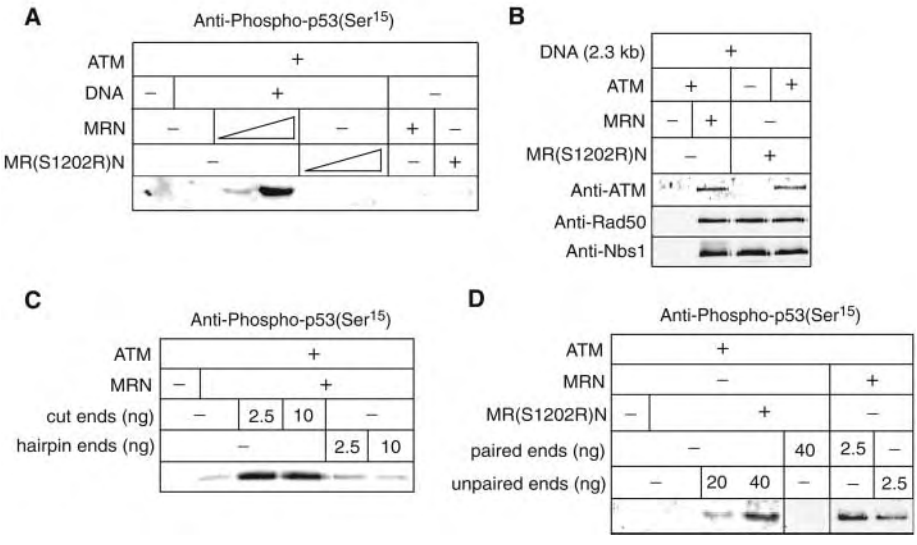


Fig. 4. Requirement of Rad50 ATP binding and DNA unwinding for DNA-dependent stimulation of ATM. (**A**) Kinase assays as in Fig. 1D with MR(S1202R)N complexes or wild-type MRN assayed with GST-p53. (**B**) DNA binding assays as in Fig. 2C, except with MR(S1202R)N mutant complexes compared with wild-type MRN. (**C**) Kinase assays with MRN, ATM, and DNA containing either normal cut ends or closed hairpin ends. (**D**) Kinase assays comparing 2.7-kb linear DNA (paired ends) with 2.7-kb DNA containing 60 bp of heterology on one end (unpaired ends).

added MRN and DNA, isolated the beads again, and analyzed the supernatant to look for dissociated ATM proteins (Fig. 3C). Flag-ATM was not observed in the supernatant because it was still bound to the beads, but one-third of the total HA-ATM used in the reaction was found in the supernatant when MRN was added to the dimeric ATM. This dissociation of HA-tagged ATM from the Flag-tagged ATM on the beads was not dependent on DNA and occurred similarly with S1981A dimeric ATM (Fig. 3D). Neither Mre11 alone nor a nonspecific protein had any effect on ATM dimer dissociation (fig. S2).

The Mre11 protein exhibits manganese-dependent nuclease activity in vitro (*12*). However, the nuclease activity of Mre11 is not active in the kinase assays shown here, because all of the reactions were performed in magnesium only, conditions which do not allow Mre11 nuclease activity (*12*) but support ATM activity in the presence of MRN.

The MRN complex also exhibits DNA binding and DNA unwinding activities that are dependent on both adenine nucleotides and Nbs1 (*13*, *14*). We tested the MR(S1202R)N mutant complex (where Ser^{1202} of Rad50 is replaced by Arg), which is specifically deficient in the adenosine triphosphate (ATP)–dependent functions of MRN (*15*), for stimulation of ATM activity. This mutant complex did not stimulate ATM activity (Fig. 4A), which indicates that at least one of the ATP-dependent activities of Rad50 is required for ATM stimulation. The deficiency of the mutant complex does not seem to be in DNA binding, because the MR(S1202R)N complex bound to DNA in equal amounts as wild-type MRN and also recruited ATM to DNA (Fig. 4B). The mutant complex also dissociated the ATM dimer similarly to the wild-type protein (fig. S2).

To test the importance of DNA unwinding for ATM stimulation, we prepared a DNA substrate with closed hairpins on each end. This substrate did not stimulate ATM activity, which indicates that opening of the DNA helix is required for MRN stimulation of dimeric ATM (Fig. 4C). We then prepared a substrate containing 60 noncomplementary base pairs at one end to mimic an unwound DNA molecule. This DNA ("unpaired ends") complemented the MR(S1202R)N mutant for ATM stimulation (Fig. 4D); thus, the MRN-specific role of ATP in this reaction may be to stimulate DNA unwinding.

In this study, we reconstituted DNA damage signaling to ATM with recombinant purified components. The accumulating evidence indicates that DNA breaks are sensed directly by the MRN complex, which binds DNA, unwinds the ends, recruits ATM, and dissociates the ATM dimer. These results are consistent with recent studies in budding yeast, which show that the Mre11-Rad50-Xrs2 complex localizes to DNA breaks very rapidly in vivo (*16*, *17*) and recruits the ATM homolog Tel1 to DNA breaks (*18*). However, our experiments do not recapitulate the requirement for ATM autophosphorylation that is observed in human cells (*9*). Effects of other factors, including protein phosphatases 5 (*19*) and 2A (*20*) and chromatin remodeling complexes, may be needed to reconstitute the effects of autophosphorylation that are seen in vivo. Our results in this study define in mechanistic detail the pathway for ATM activation after the formation of a DNA double-strand break and provide a biochemical foundation for the characterization of other factors that influence the activity of ATM in cells.

References and Notes

1. T. H. Stracker, J. W. Theunissen, M. Morales, J. H. Petrini, *DNA Repair (Amst.)* **3**, 845 (2004).
2. T. Uziel *et al.*, *EMBO J.* **22**, 5612 (2003).
3. G. Buscemi *et al.*, *Mol. Cell. Biol.* **21**, 5214 (2001).
4. P. M. Girard, E. Riballo, A. C. Begg, A. Waugh, P. A. Jeggo, *Oncogene* **21**, 4191 (2002).

5. J. H. Lee *et al.*, *Mol. Cancer Res.* **1**, 674 (2003).
6. R. Kitagawa, C. J. Bakkenist, P. J. McKinnon, M. B. Kastan, *Genes Dev.* **18**, 1423 (2004).
7. P. T. Yazdi *et al.*, *Genes Dev.* **16**, 571 (2002).
8. J. H. Lee, T. T. Paull, *Science* **304**, 93 (2004).
9. C. J. Bakkenist, M. B. Kastan, *Nature* **421**, 499 (2003).
10. Materials and methods are available as supporting material on *Science* Online.
11. J. H. Lee, T. T. Paull, data not shown.
12. T. T. Paull, M. Gellert, *Mol. Cell* **1**, 969 (1998).
13. J.-H. Lee *et al.*, *J. Biol. Chem.* **278**, 45171 (2003).
14. T. T. Paull, M. Gellert, *Genes Dev.* **13**, 1276 (1999).
15. G. Moncalian *et al.*, *J. Mol. Biol.* **335**, 937 (2004).
16. R. Shroff *et al.*, *Curr. Biol.* **14**, 1703 (2004).
17. M. Lisby, J. H. Barlow, R. C. Burgess, R. Rothstein, *Cell* **118**, 699 (2004).
18. D. Nakada, K. Matsumoto, K. Sugimoto, *Genes Dev.* **17**, 1957 (2003).
19. A. Ali *et al.*, *Genes Dev.* **18**, 249 (2004).
20. A. A. Goodarzi *et al.*, *EMBO J.* **23**, 4451 (2004).
21. Molecular interaction data have been deposited in the Biomolecular Interaction Network Database (BIND) with accession codes 216020 to 216045. We thank M. Kastan and R. Abraham for expression constructs; D. Ramsden, M. Gellert, and M. O'Dea for Rag1/Rag2 protein; S. Stevens for technical advice; members of the Paull lab for their help; and R. Rothstein for a helpful word. This work was supported by NIH (grant CA094008) and by the American Cancer Society (grant RSG-04-173-01-CCG).

Supporting Online Material
www.sciencemag.org/cgi/content/full/1108297/DC1
Materials and Methods
Figs. S1 and S2
References

6 December 2004; accepted 24 February 2005
Published online 24 March 2005;
10.1126/science.1108297
Include this information when citing this paper.

Comparative Metagenomics of Microbial Communities

Susannah Green Tringe,[1,2]* Christian von Mering,[3]* Arthur Kobayashi,[1] Asaf A. Salamov,[1] Kevin Chen,[4] Hwai W. Chang,[5] Mircea Podar,[5] Jay M. Short,[5] Eric J. Mathur,[5] John C. Detter,[1] Peer Bork,[3] Philip Hugenholtz,[1] Edward M. Rubin[1,2]†

The species complexity of microbial communities and challenges in culturing representative isolates make it difficult to obtain assembled genomes. Here we characterize and compare the metabolic capabilities of terrestrial and marine microbial communities using largely unassembled sequence data obtained by shotgun sequencing DNA isolated from the various environments. Quantitative gene content analysis reveals habitat-specific fingerprints that reflect known characteristics of the sampled environments. The identification of environment-specific genes through a gene-centric comparative analysis presents new opportunities for interpreting and diagnosing environments.

Despite their ubiquity, relatively little is known about the majority of environmental microorganisms, largely because of their resistance to culture under standard laboratory conditions. A variety of environmental sequencing projects targeted at 16*S* ribosomal RNA (rRNA) (*1*, *2*) has offered a glimpse into the phylogenetic diversity of uncultured organisms. The direct sequencing of environmental samples has provided further valuable insight into the lifestyles and metabolic capabilities of uncultured organisms occupying various environmental niches. The latter efforts include the sequencing of individual large-insert bacterial artificial chromosome (BAC) clones as well as small-insert libraries made directly from environmental DNA (*3*–*7*). The application of high-throughput shotgun sequencing environmental samples has recently provided global views of those communities not obtainable from 16*S* rRNA or BAC clone–sequencing surveys (*6*, *7*). The sequence data have also posed challenges to genome assembly, which suggests that complex communities will demand enormous sequencing expenditure for the assembly of even the most predominant members (*7*).

A practical question emerging from environmental sequencing projects is the extent to which the data are interpretable in the absence of significant individual genome assemblies. Most microbial communities are extremely complex and thus not amenable to genome assembly (*8*). This obstacle may in part be offset by the high gene density of prokaryotes [~1 open reading frame per 1000 base pairs (bp)] and currently attainable read lengths (700 to 750 bp), which result in most individual sequences containing a significant portion of at

[1]Department of Energy (DOE) Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA. [2]Lawrence Berkeley National Laboratory, Genomics Division, Berkeley, CA 94720, USA. [3]European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany. [4]University of California, Berkeley, Department of Electrical Engineering and Computer Science, Berkeley, CA 94720, USA. [5]Diversa Corporation, 4955 Directors Place, San Diego, CA 92121, USA.

*S.G.T. and C.v.M. contributed equally to this work.
†To whom correspondence should be addressed. E-mail: emrubin@lbl.gov

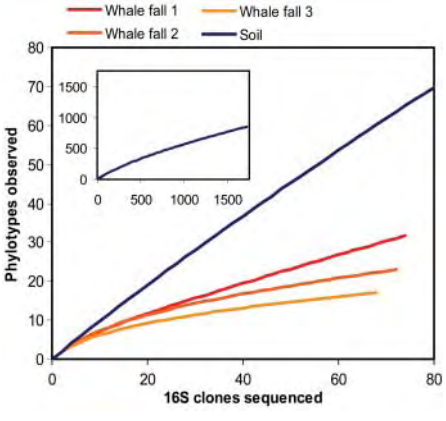


Fig. 1. Species complexity. Rarefaction curves of bacterial 16*S* rRNA clone sequences for soil and whale fall samples. (**Inset**) Rarefaction curve for all 1700 soil clones. The three whale falls are: 1, Santa Cruz Basin bone; 2, Santa Cruz Basin microbial mat; and 3, Antarctic bone.

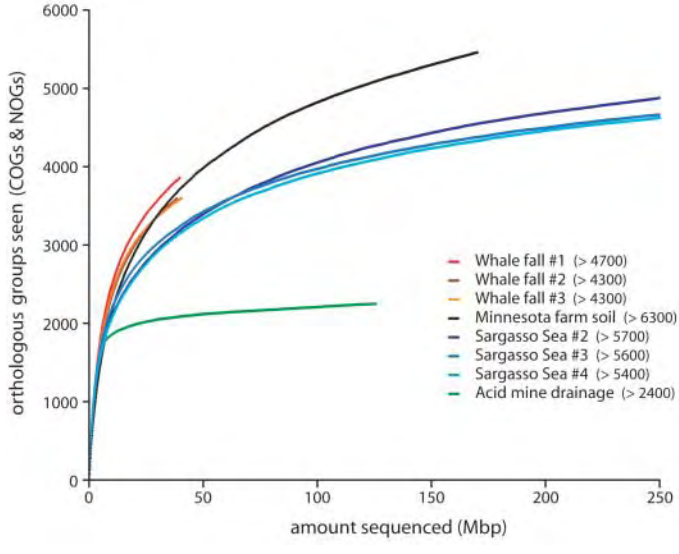


Fig. 2. Identification of orthologous groups with greater sequencing depth. The number of orthologous groups observed at least once is shown as a function of the raw sequence generated. Numbers in parentheses indicate lower limits of the total number of groups in the sample.

least one gene (*9*). Accordingly, although microbial as well as animal sequencing studies have typically targeted complete genomes, for metagenomic data, this approach may not always be necessary or feasible. Determining the proteins encoded by a community, rather than the types of organisms producing them, suggests a means to distinguish samples on the basis of the functions selected for by the local environment and reveals insights into features of that environment. In the present study, we took a gene-centric approach to environmental sequencing in our analysis of several disparate microbial communities.

The samples we characterized were derived from agricultural soil and from three isolated deep-sea "whale fall" carcasses (*10*). In contrast to the nutrient-poor environments previously subjected to large-scale metagenomic sequencing (*6*, *7*), each of these environments was nutrient-rich, albeit with very different nutrient sources (plant material for soil and lipid-rich bone for deep-sea whale fall samples). We first analyzed the microbial diversity in these samples through polymerase chain reaction (PCR)–amplified small rRNA libraries generated for each sample by using primers specific for Bacteria, Archaea, and Eukaryota. In the soil sample, a wide diversity of bacteria, few archaeal species, and some fungi and unicellular eukaryotes were found (fig. S2). We sequenced a total of 1700 clones from two independent libraries of PCR-amplified bacterial 16*S* rRNA sequences prepared from the soil DNA, and we identified at least 847 distinct ribotypes from more than a dozen phyla (fig. S2B). A rarefaction curve built from these data failed to reach saturation, and coverage estimators such as Chao1 (*11*, *12*) predicted the total number of bacterial ribotypes in this sample to be more than 3000 (Fig. 1; fig. S1), which reflected the enormous diversity found in soil (*8*). The most common ribotype accounts for 112 (6.6%) of the clones (fig. S2A) when a 97% identity cutoff is used and 81 (4.8%) when 98% identity is required. The whale fall samples are both less diverse and less evenly distributed than the soil cohort and are estimated to contain between 25 and 150 distinct ribotypes of which the most abundant accounts for 15 to 25% of the library (Fig. 1; fig. S4). The reduced species and phyla diversity of the whale fall microbial communities as compared with soil is consistent with the extreme and specialized nature of this deep-ocean ecological niche.

We explored the genomic diversity of the communities by sequencing genomic small-insert libraries made from all four samples. In light of the organismal complexity seen in the soil sample, we generated 100 million bp (Mbp) of sequence from this sample and 25 Mbp for each whale fall library. Consistent with the predicted high species diversity in the soil sample, attempts at sequence assembly were largely unsuccessful. Less than 1% of the nearly 150,000 reads generated from the soil library exhibited overlap with reads from independent clones. On the basis of our 16*S* rRNA data and the overlaps in the genomic sequence, we projected that somewhere between two and five billion base pairs of sequence would be necessary to obtain the eightfold coverage traditionally targeted for draft genome assemblies, even for the single most predominant genome in this complex community (*13*). For each whale fall library, we estimate that between 100 and 700 Mbp of shotgun sequence data would be needed in order to generate a draft assembly for the most prevalent genome. Assembling genomes for low-abundance community members in any of these samples would clearly require significantly more sequence data.

Given these hurdles to the assembly of complete genomes from the samples, we investigated the genes present without attempting to place them in the context of an individual genome. In preliminary studies, we compared gene predictions from assembled sequence with unassembled, using available metagenomic data (*13*). With our analysis supporting the validity of gene predictions on unassembled reads, we applied an automated annotation process to the sequence data from several different environmental samples. As our analysis relied primarily on the predicted genes on small DNA fragments, the majority of which were individual sequence reads, we termed each environmental sequence an environmental gene tag (EGT), to distinguish EGTs from the sequencing reads primarily used for the assembly of genomes. The gene contents of the partially

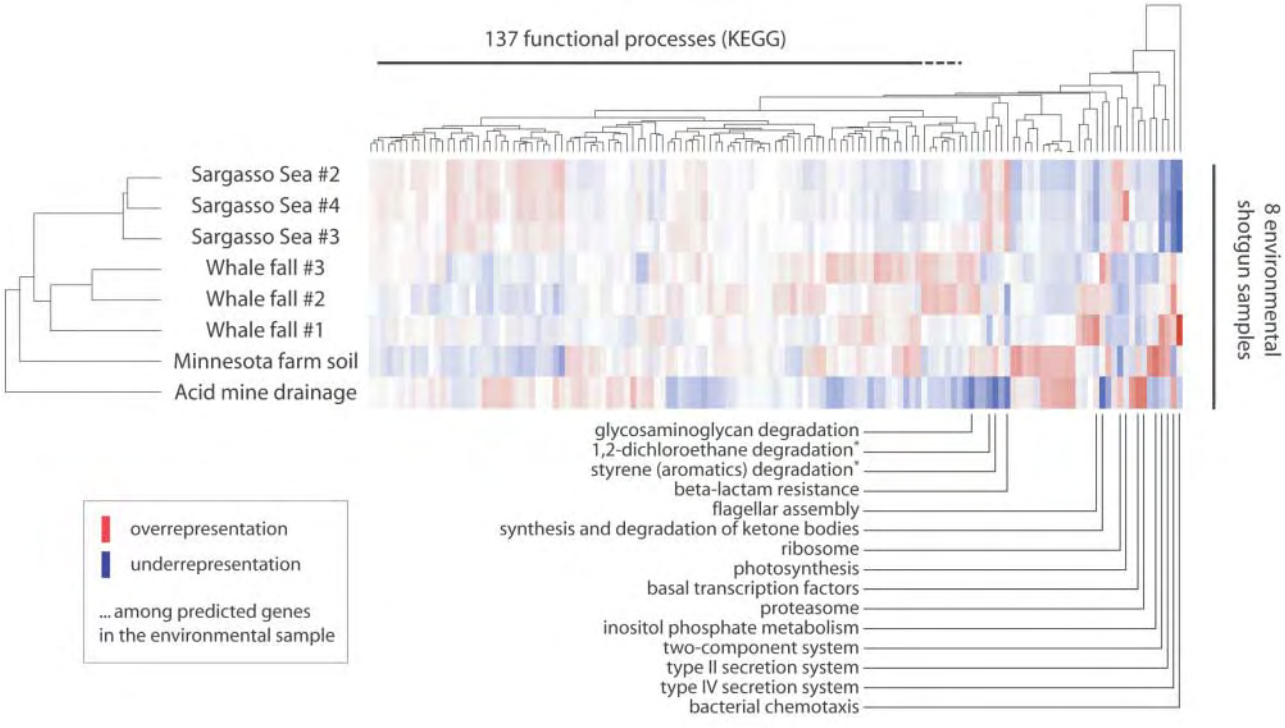


Fig. 3. Functional profiling of microbial communities. Two-way clustering of samples and encoded functions based on relative enrichment of KEGG functional processes. The 15 most discriminating processes are highlighted. Asterisks indicate that environmental genes mapping to these processes probably have a broader range of substrates than the KEGG process title indicates.

assembled and unassembled reads from soil and whale fall samples were compared with each other and with those of an acid mine drainage biofilm community (*6*) and with each of three independent samples from Sargasso Sea surface waters (*7*). Putative genes were predicted on at least 90% of the EGTs from all samples, even when the sequence fragments were individual reads. More than a third of the EGTs contained two or more predicted open reading frames, which raised the possibility of nearest-neighbor analysis (*14*).

Roughly half of the predicted proteins in each sample showed homology to orthologous groups in an expanded in-house COG (clusters of orthologous groups of proteins) database (*15*, *16*). To test whether the orthologous groups observed in a limited sampling of each library were representative of the full range of groups in a community, we plotted the number of orthologous groups detected at increasing levels of sequencing depth. For all samples, saturation for frequently occurring orthologous groups is observed after a modest amount of sequencing, whereas the general slope of the curve reveals information about community diversity (Fig. 2). In the relatively simple acid mine drainage biofilm community, 90% of the orthologous groups were detected with just 25 Mbp of raw sequence (~15 Mbp quality sequence)—a fraction of that needed to assemble genomes. Even in the considerably more complex soil community, the curve starts to flatten at 25 Mbp, which suggests that new orthologous groups detected at this point are found only in a minority of the community members. The Sargasso Sea communities, consistent with their species complexity, fell between acid mine drainage and soil; the whale falls, however, exhibited trajectories quite similar to soil. We observed qualitatively similar curves when limiting the analysis to the 4873 COGs contained in the 2003 release or to the domain-oriented Pfam database (*17*) (fig. S5), which suggests that this phenomenon is not an artifact of comparison to a particular database.

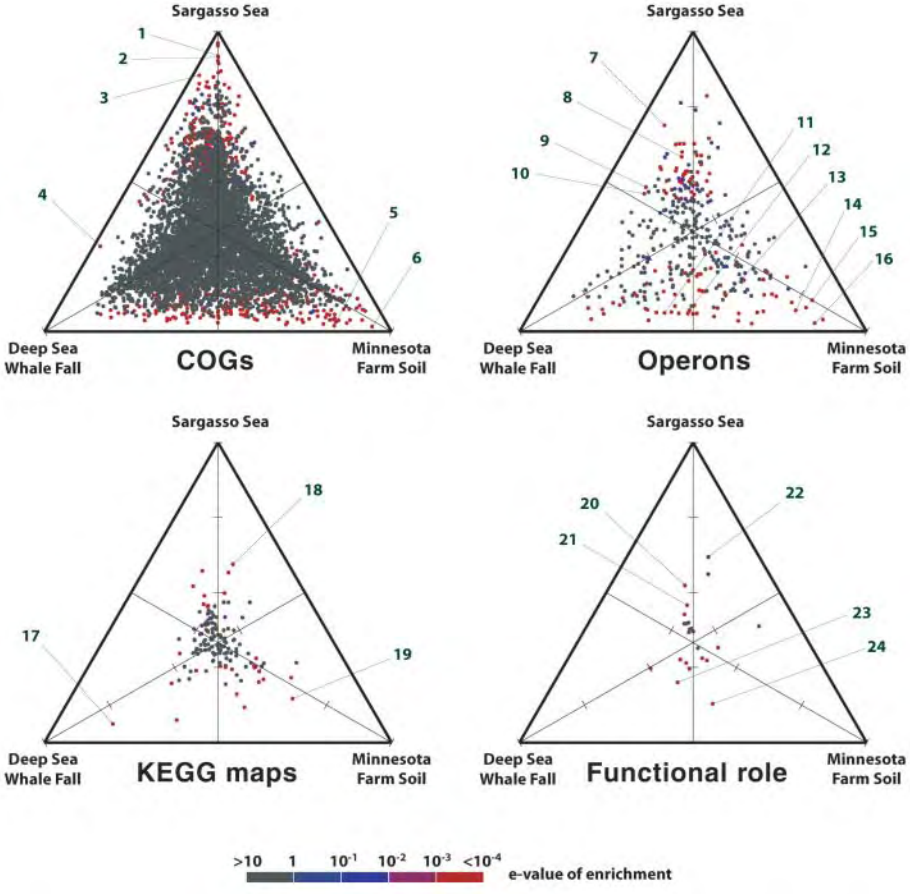


Fig. 4. Specific enrichments. Three-way comparisons of soil, whale fall, and Sargasso Sea environments in terms of COGs, operons, KEGG processes, and COG functional categories. Each dot shows the relative abundance of an item in the three environmental samples, such that proximity to a vertex is proportional to the level of enrichment in the respective sample. Color indicates statistical significance of the enrichment. Marked items discussed in main text: 1, COG5524 bacteriorhodopsin; 5, COG3459 cellobiose phosphorylase; 7, ABC-type proline/glycine betaine transport system; 10, Na^+-transporting NADH:ubiquinone reductase; 14, osmosensitive, active K^+-transport system; 18, photosynthesis; and 19, type I polyketide biosynthesis (antibiotics). A complete listing of numbered items is available in the SOM, and an enhanced version of the figure is at (*23*).

We next explored the relative proportion of the total protein sets devoted to particular functions in a sample, given evidence that not only message levels (*18*), but also library representation (*19*), of genes coding for specialized enzymes can vary with sample source. We specifically explored whether independent samples from similar, although geographically separated, environments would exhibit functional profiles more similar to each other than to those from disparate environments. We binned predicted proteins into functional categories at four levels: first, individual genes (orthologous groups inferred from sequenced genomes); second, groups of genes frequently observed as neighbors in complete genomes ["operons," shown to correlate with metabolic and other pathways (*20*)]; third, higher order cellular processes from the manually curated KEGG (Kyoto Encyclopedia of Genes and Genomes) database (*21*); and fourth, broad functional categories from the COG database (*13*, *15*). Assembled contigs were weighted to account for the number of independent clones contributing to them.

A two-way clustering of samples and KEGG maps, in which over- and underrepresented categories are indicated by red and blue blocks, respectively, is displayed in Fig. 3 (fig. S6 shows similar figures based on COGs and operons). Regardless of the functional binning employed, the independent Sargasso Sea samples clustered together, as did the whale fall samples. These profiles clearly suggest that the predicted protein complement of a community is similar to that of other communities whose environments of origin pose similar metabolic demands. Our results further support the hypothesis that the "functional" profile of a community is influenced by its environment and that EGT data can be used to develop fingerprints for particular environments.

To assess the significance of these similarities and differences and to identify functions of importance for communities existing in specific environments, we systematically examined the differences in gene content between samples (Fig. 4). For this analysis, the three whale fall samples were pooled together, as were the three ocean samples. At each level, significant differences among the respective microbial communities were observed that suggested environment-specific variations in both biochemistry and phylogeny. The acid mine drainage was not included in this analysis because of its great dissimilarity from the other samples (Fig. 3; fig. S6) and low species diversity, both likely reflective of the very extreme nature of this environment.

At the individual gene level, quite a few orthologous groups are exclusive to a particular environment (Fig. 4, upper left). For example, 73 putative orthologs of cellobiose phosphorylase, involved in degradation of plant material, are found in the ~100 Mbp of soil sequence, but none are found in the ~700 Mbp of sequence examined from the Sargasso Sea. On the other

hand, 466 distinct homologs of the light-driven proton pump bacteriorhodopsin are found in the surface waters of the Sargasso Sea, whereas none are found in the deep-sea whale falls or in soil.

The analysis of operons likewise reveals similarities and differences in functional systems (Fig. 4, upper right) that suggest features of the environments. The most discriminating operons tend to be systems for the transport of ions and inorganic components, highlighting their importance for survival and adaptation. With respect to ionic and osmotic homeostasis, for example, the two maritime environments are similar—both show a strong enrichment in operons that contain transporters for organic osmolites and sodium ion exporters coupled to oxidative phosphorylation. The soil sample, on the other hand, has a strong enrichment in operons responsible for active potassium channeling. These biases nicely reflect the relative abundance of these ions in the respective environments: Whereas typical ocean water contains considerably more sodium ions than potassium, the soil sample examined here contained high potassium and low sodium concentrations (*13*).

Examination of higher order processes reveals known differences in energy production (e.g., photosynthesis in the oligotrophic waters of the Sargasso Sea and starch and sucrose metabolism in soil) (*7*) or population density and interspecies communication [overrepresentation of conjugation systems, plasmids, and antibiotic biosynthesis in soil (Fig. 4, lower left)] (*22*). The broad functional COG categories, on the other hand, primarily suggest differences in genome size and phylogenetic composition (*13*).

Notably, many uncharacterized genes and processes are among the most overrepresented categories in each sample. This hints at an abundance of previously unknown functional systems, specific to each environment, whose occurrence patterns may offer useful guidance for further, more directed experimental and computational investigations. More extensive sampling in both time and space will reveal which features are broadly distributed within a given environment and which are unique to the places and times sampled here. Nonetheless, this analysis of genes and functional modules in environments reveals expected contrasts, hints at certain nutrition conditions, and points to novel genes and systems contributing to a particular "life-style" or environmental interaction.

The predicted metaproteome, based on fragmented sequence data, is sufficient to identify functional fingerprints that can provide insight into the environments from which microbial communities originate. Information derived from extension of the comparative metagenomic analyses performed here could be used to predict features of the sampled environments such as energy sources or even pollution levels. At the same time, the environment-specific distribution of unknown orthologous groups and operons offers exciting avenues for further investigation. Just as the incomplete but information-dense data represented by expressed sequence tags have provided useful insights into various organisms and cell types, EGT-based ecogenomic surveys represent a practical and uniquely informative means for understanding microbial communities and their environments.

References and Notes

1. E. F. DeLong, N. R. Pace, *Syst. Biol.* **50**, 470 (2001).
2. P. Hugenholtz, *Genome Biol* **3**, REVIEWS0003 (2002).
3. M. R. Liles, B. F. Manske, S. B. Bintrim, J. Handelsman, R. M. Goodman, *Appl. Environ. Microbiol.* **69**, 2684 (2003).
4. O. Beja *et al.*, *Science* **289**, 1902 (2000).
5. A. H. Treusch *et al.*, *Appl. Environ. Microbiol.* **6**, 970 (2004).
6. G. W. Tyson *et al.*, *Nature* **428**, 37 (2004).
7. J. C. Venter *et al.*, *Science* **304**, 66 (2004).
8. V. Torsvik, L. Ovreas, T. F. Thingstad, *Science* **296**, 1064 (2002).
9. Y. A. Goo *et al.*, *BMC Genomics* **5**, 3 (2004).
10. C. R. Smith, A. R. Baco, in *Oceanography and Marine Biology: An Annual Review*, R. N. Gibson, R. J. A. Atkinson, Eds. (Taylor & Francis, London, 2003), vol. 41, pp. 311–354.
11. J. B. Hughes, J. J. Hellmann, T. H. Ricketts, B. J. Bohannan, *Appl. Environ. Microbiol.* **67**, 4399 (2001).
12. R. K. Colwell, personal communications (1994–2004).
13. Supplementary online material.
14. R. Overbeek, M. Fonstein, M. D'Souza, G. D. Pusch, N. Maltsev, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2896 (1999).
15. R. L. Tatusov *et al.*, *BMC Bioinformatics* **4**, 41 (2003).
16. C. von Mering *et al.*, *Nucleic Acids Res.* **33**, D433 (2005).
17. A. Bateman *et al.*, *Nucleic Acids Res.* **32** (Database special issue), D138 (2004).
18. S. K. Rhee *et al.*, *Appl. Environ. Microbiol.* **70**, 4303 (2004).
19. D. E. Robertson *et al.*, *Appl. Environ. Microbiol.* **70**, 2429 (2004).
20. C. von Mering *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15428 (2003).
21. M. Kanehisa, S. Goto, S. Kawashima, Y. Okuno, M. Hattori, *Nucleic Acids Res.* **32** (Database special issue), D277 (2004).
22. R. Daniel, *Curr. Opin. Biotechnol.* **15**, 199 (2004).
23. http://string.embl.de/metagenome_comp_suppl/

24. This work was performed under the auspices of the DOE's Office of Science, Biological and Environmental Research Program; the University of California, Lawrence Livermore National Laboratory, under contract no. W-7405-Eng-48; Lawrence Berkeley National Laboratory under contract no. DE-AC03-76SF00098; and Los Alamos National Laboratory under contract no. W-7405-ENG-36. S.G.T. was supported by grant no. THL007279F, an NIH National Research Service Award (NRSA) Training and Fellowship grant to E.R. K.C. was supported by NSF grant no. EF 03-31494. Sequencing of the environmental libraries was performed under a license agreement with Diversa (J. R. Short, U.S. patent no. 6455254). We gratefully acknowledge the efforts of C. Baptista, L. Christoffersen, J. Garcia, K. Li, J. Ritter, P. Sammon, S. Wells, D. Whitney, J. Eads, T. Richardson, M. Noordewier, and L. Bibbs. We thank C. Smith for providing the whale fall samples; K. Remington for providing Sargasso Sea sample information; N. Ivanova, N. Kyrpides, and members of the Rubin laboratory for helpful comments on the manuscript; and J. Chapman, I. Grigoriev, E. Szeto, J. Korbel, T. Doerks, K. Foerstner, E. Harrington, and M. Krupp for assistance with data processing and analysis. These Whole Genome Shotgun projects have been deposited with the DNA Data Bank of Japan, the European Molecular Biology Laboratory (EMBL) Nucleotide Sequence Database, and the GenBank in collaboration (DDBJ/EMBL/GenBank) under the project accessions AAFX00000000 (soil), AAFY00000000 (whale fall 1), AAFZ00000000 (whale fall 2), and AAGA00000000 (whale fall 3). For each project, the version described in this paper is the first version, AAFX01000000, AAFY01000000, AAFZ01000000, and AAGA01000000. The 16*S* rRNA sequences from the soil and three whale fall samples have been deposited under GenBank accession nos. AY921654 to AY922252. The metagenomic data will also be incorporated into the U.S. Department of Energy Joint Genome Institute Integrated Microbial Genomes system (www.jgi.doe.gov/) to facilitate detailed comparative analysis of the data in the context of all publicly available complete microbial genomes.

Supporting Online Material
www.sciencemag.org/cgi/content/full/308/5721/554/DC1
Materials and Methods
Figs. S1 to S7
References and Notes

23 November 2004; accepted 4 February 2005
10.1126/science.1107851

A Cellular MicroRNA Mediates Antiviral Defense in Human Cells

Charles-Henri Lecellier,[1]* Patrice Dunoyer,[1] Khalil Arar,[2] Jacqueline Lehmann-Che,[3] Stephanie Eyquem,[4] Christophe Himber,[1] Ali Saïb,[3] Olivier Voinnet[1]*

In eukaryotes, 21- to 24-nucleotide-long RNAs engage in sequence-specific interactions that inhibit gene expression by RNA silencing. This process has regulatory roles involving microRNAs and, in plants and insects, it also forms the basis of a defense mechanism directed by small interfering RNAs that derive from replicative or integrated viral genomes. We show that a cellular microRNA effectively restricts the accumulation of the retrovirus primate foamy virus type 1 (PFV-1) in human cells. PFV-1 also encodes a protein, Tas, that suppresses microRNA-directed functions in mammalian cells and displays cross-kingdom antisilencing activities. Therefore, through fortuitous recognition of foreign nucleic acids, cellular microRNAs have direct antiviral effects in addition to their regulatory functions.

In plants and insects, viral double-stranded RNA is processed into small interfering RNAs (siRNAs) by the ribonuclease (RNase) III enzyme Dicer. These siRNAs are incorporated into the RNA-induced silencing complex to target the pathogen's genome for destruction (*1*, *2*). Plant and insect viruses can counter this defense with silencing suppres-

sor proteins, which often have adverse side effects on microRNA (miRNA) functions (*3*, *4*). Although undisputed in plants and insects, a defensive role for RNA silencing in vertebrates has not been demonstrated. Virus-derived small RNAs have not been detected in infected vertebrate cells, with the exception of miRNAs produced by the Epstein-Barr virus, but the role of those molecules remains unclear (*5*). Moreover, some mammalian virus-encoded proteins that suppress RNA silencing have only been investigated in heterologous systems (*6*). Because RNA silencing suppresses mobilization of endogenous retroviruses in plants, yeast, worms, and flies (*7*), we reasoned that retrotransposition of mammalian exogenous viruses might also be subject to this process. Therefore, we studied the primate foamy virus type 1 (PFV-1), a complex retrovirus (akin to human immunodeficiency virus) that, in addition to the Gag, Pol, and Env proteins, produces two auxiliary factors, Bet and Tas, from the internal promoter (IP) (Fig. 1A) (*8*).

PFV-1 accumulation was strongly enhanced in 293T cells expressing the P19 silencing suppressor (Fig. 1B). This suggested that a siRNA and/or miRNA pathway limits PFV-1 replication in human cells, because P19 specifically binds to and inactivates both types of small RNAs (*4*, *9*, *10*). Viral sequences spanning the 12-kb-long PFV-1 genome (Fig. 1A) were fused to the 3′ untranslated region (UTR) of a green fluorescent protein (GFP)–tagged reporter gene,

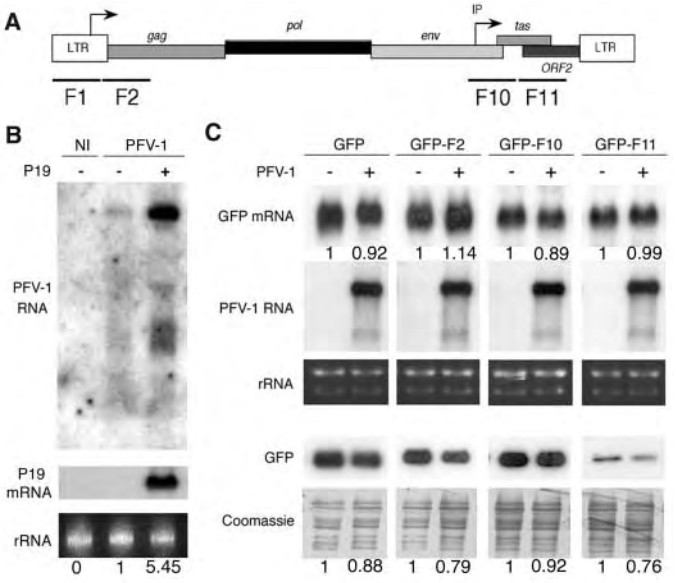


Fig. 1. RNA silencing limits PFV-1 accumulation in mammalian cells. (**A**) Schematic of the PFV-1 genome. Bent arrows indicate the start of transcription between the 5′-proximal long-terminal repeat (LTR) and the IP. Viral sequences (F1 to F10) used for GFP transcriptional fusions are indicated. (**B**) mRNA accumulation from PFV-1 in 293T cells that do (+) or do not (–) stably express the P19 protein. Cells were harvested 48 hours after transfection. Northern analysis confirms P19 expression. rRNA, ethidium bromide staining of ribosomal RNA; NI, noninfected. (**C**) The GFP sensors F1 to F11 were transfected together with (+) or in the absence of (–) PFV-1. Their expression was assayed 48 hours later by Northern (first upper panel) and Western (fourth panel) analysis. (Second upper panel) PFV-1 RNA accumulation. (Bottom) Staining of total protein for loading control. Relative RNA or protein accumulation is shown at the bottom of each panel, with control levels arbitrarily set to 1.

[1]CNRS Unité Propre de Recherche (UPR) 2357, Institut de Biologie Moléculaire des Plantes, 12 rue du Général Zimmer, 67084 Strasbourg Cedex, France. [2]Proligo, Paris, France. [3]CNRS UPR9051, Hôpital St-Louis, Paris, France. [4]INSERM U462, Hôpital St-Louis, Paris, France.

*To whom correspondence should be addressed. E-mail: charles.lecellier@infobiogen.fr (C.-H.L.); olivier.voinnet@ibmp-ulp.u-strasbg.fr (O.V.)

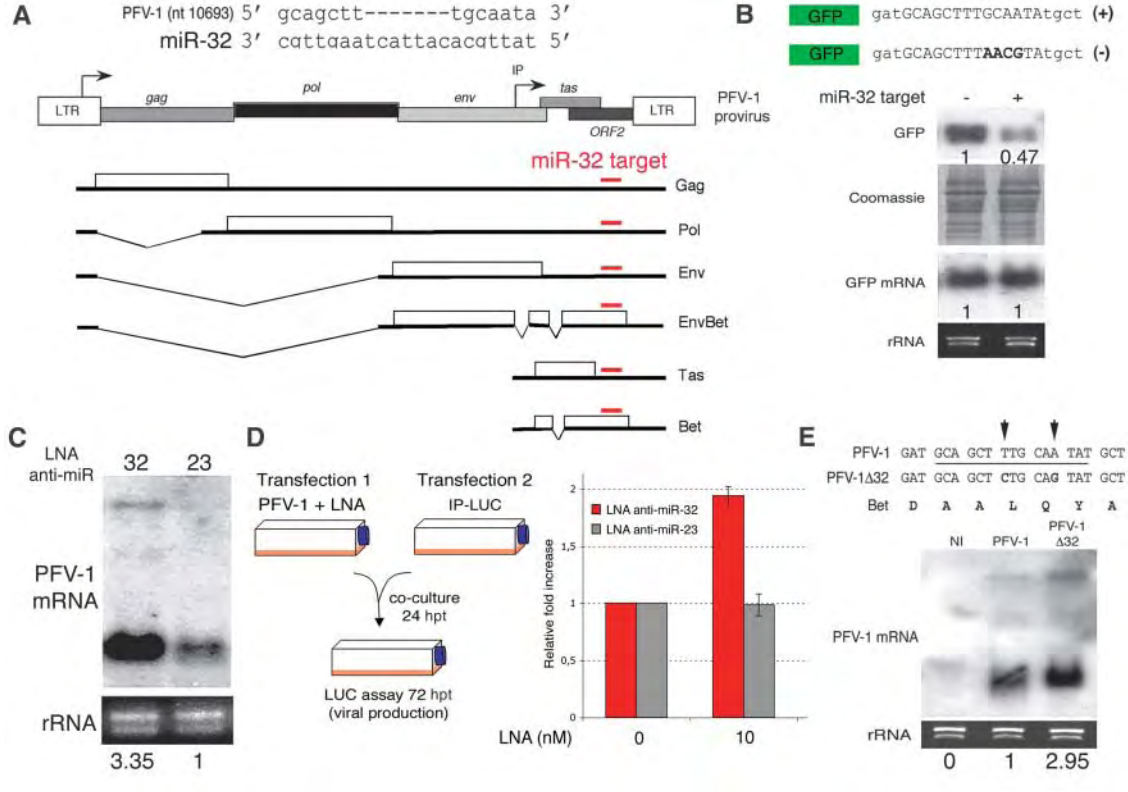


Fig. 2. miR-32 effectively limits PFV-1 replication. (**A**) Position of the computationally predicted miR-32 target relative to PFV-1 transcripts. (**B**) The miR-32 target sequence or a mutated form thereof (–) was fused to the 3′UTR of a GFP reporter gene (+). Constructs were transfected in HeLa cells and harvested 48 hours later. GFP and GFP mRNA accumulation were assessed by Western (top) and Northern (bottom) analysis. (**C**) HeLa cells were transfected with PFV-1 together with LNAs (10 nM) directed against miR-32 or miR-23. Total RNA was extracted 48 hours after transfection and subjected to Northern analysis. (**D**) PFV-1 was transfected in HeLa cells (transfection 1). Separate cells were transfected with a luciferase-based reporter (Luc) driven by the PFV-1 IP, which is activated by the transactivator Tas (transfection 2). Transfections 1 and 2 were mixed 24 hours later and further cocultured for 48 hours. Luciferase expression in cells from transfection 2, resulting from their infection by virions released from transfection 1, was then quantified. hpt, hours post-transfection. (**E**) The miR-32 target sequence within PFV-1Δ32 contains two synonymous mutations (arrows). Northern analysis of mutant and wild-type virus mRNAs was carried out 48 hours after transfection.

and the resulting constructs (F1 to F11) were cotransfected with PFV-1 into baby hamster kidney (BHK) 21 cells. Any viral-derived siRNA would induce RNA silencing of the corresponding reporter fusions, diagnosed as reduced GFP mRNA accumulation. However, the mRNA levels from those constructs were similar in noninfected and infected cells (Fig. 1C). Use of a highly sensitive RNase protection assay likewise failed to provide evidence for viral-derived siRNAs (fig. S1).

The GFP levels from fusion F11 were disproportionably reduced compared to the accumulation of the F11 mRNA (Fig. 1C). They were also reduced compared to the GFP levels from constructs F2 and F10. Although a possible result of intrinsic protein instability, the effect was reminiscent of the translational inhibition directed by animal miRNAs (*11*). However, it was independent of the presence or absence of PFV-1 (Fig. 1C), suggesting that any miRNA involvement was likely cellular rather than viral. Using the DIANA-microT algorithm (*12*), we found a high probability hit (free energy of –21.0 kcal/mol) between the PFV-1 F11 sequence and the human miR-32 (Fig. 2A) (*13*). The predicted miR-32 target sequence was sufficient to promote translation inhibition of the GFP mRNA (Fig. 2B), unlike a derivative thereof that carried four mutations disrupting annealing of the small RNA. Moreover, translation inhibition by miR-32 was suppressed in P19-expressing cells (fig. S2).

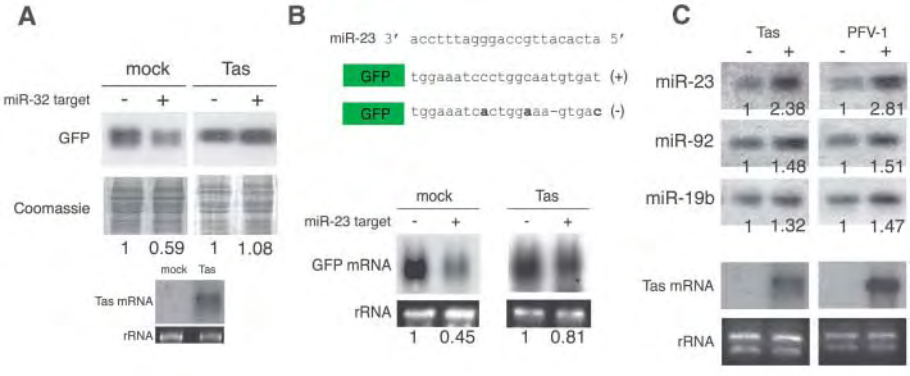


Fig. 3. Tas suppresses miRNA-directed silencing in mammalian cells. (**A**) The reporter constructs used in Fig. 2B were transected in control BHK21 cells (mock) or in cells stably expressing Tas. GFP expression was assayed by Western analysis (top) 48 hours after transfection. Tas expression was confirmed by Northern analysis (bottom). (**B**) A sequence with 100% complementarity to miR-23 (+) or a mutated derivative thereof (–) was inserted into the 3′UTR of the GFP reporter gene. Constructs were transfected in BHK21 cells (mock) or in cells stably expressing Tas (Tas), and the GFP mRNA was assayed by Northern analysis 48 hours later. (**C**) Northern analysis of cellular miRNAs from BHK21 cells expressing (+) or not expressing (–) Tas (left) and from noninfected (–) or PFV-1–infected (+) BHK21 cells (right). Total RNA was extracted 5 days after infection.

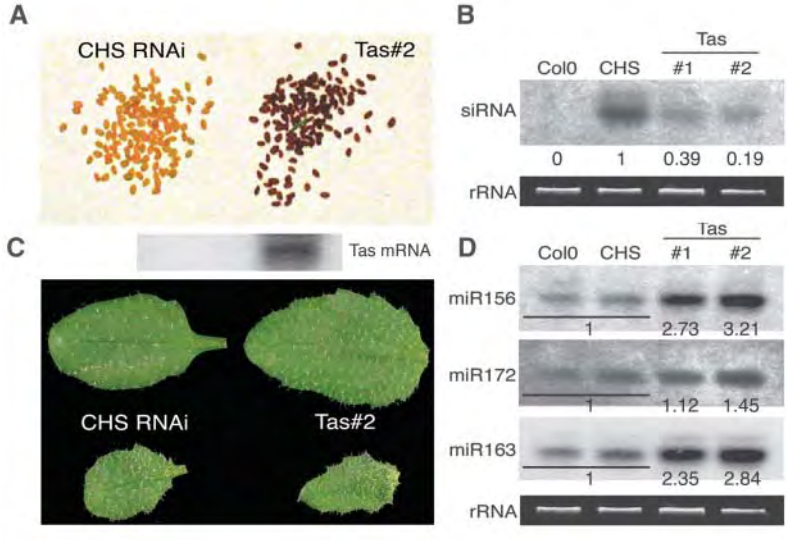


Fig. 4. (**A**) Transgenic Tas suppresses CHS RNAi in *Arabidopsis*. (**B**) Northern analysis of CHS siRNAs in two independent Tas-expressing lines. Col0, nontransformed plants; CHS, the reference RNAi line. (**C**) Developmental defects and (**D**) miRNA accumulation in Tas-expressing *Arabidopsis*. miR156 and miR172 are evolutionarily conserved miRNAs that promote cleavage and translation inhibition, respectively. miR163 is a cleavage-promoting, *Arabidopsis*-specific miRNA.

The miR-32 target is in open reading frame (ORF) 2, shared by the Bet and EnvBet proteins, and is also within the 3′UTR of all remaining PFV-1 mRNAs (Fig. 2A). To address the antiviral effect of miR-32, we used antisense locked nucleic acid (LNA) oligonucleotides (fig. S3), which yield highly stable hybrids (*14*). In HeLa and BHK-21 cells, the transfected anti-miR-32 LNA prevented translation inhibition by miR-32, whereas a control LNA with antisense sequence of the unrelated miR-23 did not (fig. S3). At LNA concentrations of 10 nM, accumulation of PFV-1 mRNAs was higher in the anti-miR-32–treated cells than in the anti-miR-23–treated cells (Fig. 2C). Use of a luciferase-based assay also indicated that the anti-miR-32, unlike the anti-miR-23, almost doubled progeny virus production (Fig. 2D).

Although these results are consistent with an antiviral effect of miR-32, we could not discard the possibility of an indirect action of anti-miR-32 LNA causing, for instance, ectopic expression of cellular miR-32 targets, which could in turn increase viral fitness. The miR-32 target sequence in PFV-1 was thus modified to contain two synonymous mutations that abolished the miR-32 pairing but preserved the Bet amino acid content (Fig. 2E). The mRNA levels from the miR-32–resistant virus (PFV-1Δ32) were three times as high as those from the unmodified virus, consistent with the anti-miR-32 results (Fig. 2, E and C). Therefore, miR-32 exerts a direct, sequence-specific effect against PFV-1.

Does PFV-1 encode a silencing suppressor to counter the antiviral effect of miR-32? The constitutive presence of miR-32 required that the putative suppressor be synthesized precociously, which is the case of the Tas and Bet proteins (Fig. 2A). As Bet is dispensable for productive replication, Tas appears the most likely candidate (*15*). miR-32–mediated translational inhibition was indeed suppressed in Tas-expressing BHK21 cells (Fig. 3A). This was not specific for the sequence or activity of miR-32, because Tas, like P19, also suppressed endonucleolytic cleavage of GFP sensors carrying a perfect miR-23 target (Fig. 3B and fig. S2). Probably as a consequence of its suppressor function, Tas promoted the nonspecific overaccumulation of all cellular miRNAs inspected, which we also observed 5 days after PFV-1 infection in BHK21 cells (Fig. 3C). miRNA overaccumulation is also seen with several plant viral suppressors that interfere with the miRNA pathway (*3*, *4*).

To validate the silencing suppression activity of Tas in a heterologous system, we used an *Arabidopsis* line expressing an RNA interference (RNAi) construct targeted against chalcone synthase (CHS), which is responsible for the brown seed-coat pigmentation (*4*). This line accumulates CHS siRNAs and, conse-

quently, produces pale yellow seeds (Fig. 4A, left). Transgenic Tas expression restored anthocyanin synthesis (Fig. 4A, right) because of a strong decrease in CHS siRNA levels (Fig. 4B). Tas-expressing plants also exhibited developmental anomalies, including leaf elongation and serration (Fig. 4C), reminiscent of those elicited in *Arabidopsis* by viral suppressors interfering with miRNA functions (*3*, *4*). As in mammalian cells, Tas enhanced miRNA accumulation (Fig. 4D), independently of their nature or mode of action, suggesting that it suppresses a fundamental step shared between the miRNA and siRNA pathways that is conserved from plants to mammals.

These results indicate that RNA silencing limits the replication of a mammalian virus, PFV-1, and that a cellular miRNA contributes substantially to this response. As a counterdefense, PFV-1 produces Tas, a broadly effective silencing suppressor. Because all our experiments were conducted with Tas-expressing viruses, because of the essential role of the protein for replication (*15*), the strong effect of Tas on siRNA accumulation observed in *Arabidopsis* could account for our failure to detect siRNAs in mammalian cells (fig. S1). Therefore, we do not yet rule out their implication in the antiviral response reported here.

Our findings with miR-32 and PFV-1 were in fact anticipated in plants by Llave, who pointed out several near-perfect homologies between *Arabidopsis* small RNAs and viral genomes (*16*). The chances of a match between cellular miRNAs and foreign (i.e., viral) RNAs increase proportionally with the size of sampled sequences. The extent to which cellular miRNAs will be selected to target pathogen genomes upon their initial interaction with viruses may vary. Endogenous viruses might effectively coevolve with miRNAs for defensive or developmental purposes (*17*, *18*), such that viral control might eventually constitute the sole function of some cellular miRNAs. Exogenous viruses with high mutation rates could, on the other hand, rapidly escape this miRNA interference through modification of the small RNA complementary regions (*19*).

Our results support the emerging notion that miRNAs might be broadly implicated in viral infection of mammalian cells, with either positive or negative effects on replication (*5*, *20*). They also indicate that virtually any miRNA has fortuitous antiviral potential, independently of its cellular function. Moreover, because the repertoire of expressed miRNAs likely varies from one cell type to another (*11*), this phenomenon could well explain some of the differences in viral permissivity observed between specific tissues.

Note added in proof: Recent findings indicate that a single 8-oligonucleotide seed (small RNA positions 1 to 8 from the 5′ end) is sufficient to confer strong regulation by animals miRNAs. Thus, fortuitous targeting of foreign RNAs by cellular miRNAs could be widespread (*21*, *22*).

References and Notes

1. A. J. Hamilton, D. C. Baulcombe, *Science* **286**, 950 (1999).
2. H. Li, W. X. Li, S. W. Ding, *Science* **296**, 1319 (2002).
3. K. D. Kasschau *et al.*, *Dev. Cell* **4**, 205 (2003).
4. P. Dunoyer, C. H. Lecellier, E. A. Parizotto, C. Himber, O. Voinnet, *Plant Cell* **16**, 1235 (2004).
5. S. Pfeffer *et al.*, *Science* **304**, 734 (2004).
6. W. X. Li *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 1350 (2004).
7. V. Schramke, R. Allshire, *Curr. Opin. Genet. Dev.* **14**, 174 (2004).
8. M. Heinkelein *et al.*, *EMBO J.* **19**, 3436 (2000).
9. K. Ye, L. Malinina, D. J. Patel, *Nature* **426**, 874 (2003).
10. J. M. Vargason, G. Szittya, J. Burgyan, T. M. Tanaka Hall, *Cell* **115**, 799 (2003).
11. D. P. Bartel, *Cell* **116**, 281 (2004).
12. M. Kiriakidou *et al.*, *Genes Dev.* **18**, 1165 (2004).
13. M. Lagos-Quintana, R. Rauhut, W. Lendeckel, T. Tuschl, *Science* **294**, 853 (2001).
14. J. S. Jepsen, M. D. Sorensen, J. Wengel, *Oligonucleotides* **14**, 130 (2004).
15. G. Baunach, B. Maurer, H. Hahn, M. Kranz, A. Rethwilm, *J. Virol.* **67**, 5411 (1993).
16. C. Llave, *Mol. Plant Pathol.* **5**, 361 (2004).
17. H. Seitz *et al.*, *Nat. Genet.* **34**, 261 (2003).
18. S. Mi *et al.*, *Nature* **403**, 785 (2000).
19. A. T. Das *et al.*, *J. Virol.* **78**, 2601 (2004).
20. S. Lu, B. R. Cullen, *J. Virol.* **78**, 12868 (2004).
21. J. Brennecke *et al.*, *PLoS Biol.* **15**, e85 (2005).
22. B. Lewis, C. Burge, D. Bartel, *Cell* **120**, 15 (2005).

23. We thank S. W. Ding, B. Cullen, and P. Zamore for critical reading of the manuscript and access to data; members of the Voinnet lab for discussions; and R. Wagner's team for excellent plant care. Supported by an Action Thématique Incitative sur Programme from the CNRS, the Fondation pour la Recherche Médicale, and the Université Louis Pasteur, Strasbourg.

Supporting Online Material

www.sciencemag.org/cgi/content/full/308/5721/557/DC1
Materials and Methods
Figs. S1 to S4
References and Notes

16 December 2004; accepted 8 February 2005
10.1126/science.1108784

Postsecretory Hydrolysis of Nectar Sucrose and Specialization in Ant/Plant Mutualism

M. Heil,* J. Rattke, W. Boland

Obligate *Acacia* ant plants house mutualistic ants as a defense mechanism and provide them with extrafloral nectar (EFN). Ant/plant mutualisms are widespread, but little is known about the biochemical basis of their species specificity. Despite its importance in these and other plant/animal interactions, little attention has been paid to the control of the chemical composition of nectar. We found high invertase (sucrose-cleaving) activity in *Acacia* EFN, which thus contained no sucrose. Sucrose, a disaccharide common in other EFNs, usually attracts nonsymbiotic ants. The EFN of the ant acacias was therefore unattractive to such ants. The *Pseudomyrmex* ants that are specialized to live on *Acacia* had almost no invertase activity in their digestive tracts and preferred sucrose-free EFN. Our results demonstrate postsecretory regulation of the carbohydrate composition of nectar.

Many plants produce nectar in their flowers (floral nectar) and on vegetative parts [extrafloral nectar (EFN)] to mediate their interactions with animals. The chemical composition of nectar strongly affects the identity and behavior of the attracted insects and thus the outcome of the interaction (*1*–*3*). Particularly important chemical factors include amino acid content (*4*–*6*) and the ratio and amount of the main sugars: glucose, fructose, and sucrose (*3*). However, previous studies have focused on nectar as a "standing crop," leaving open the question of how its chemical composition is controlled.

Floral nectar is produced to attract pollinators, whereas EFN acts to defend plants indirectly [see (*7*) for a description of EFN in more than 80 plant families]. Most interactions among animals and both floral and extrafloral nectars are thus believed to be mutualistic. Highly specialized mutualisms are surprisingly rare in nature, because they are associated with specific coevolutionary problems (*8*). In mutualisms in general, one partner provides a service for the other and receives some kind of reward (*9*). In defensive ant/plant mutualisms, the presence of ants serves as an indirect defense mechanism and, in return, they receive food rewards and/or nesting space (*10*).

Ant/plant mutualisms differ widely in their specificity and thus are particularly suitable for

Department of Bioorganic Chemistry, Max-Planck-Institute for Chemical Ecology, Hans-Knöll-Strasse 8, D-07745 Jena, Germany.

*To whom correspondence should be addressed at FB 9 BioGeo-Allgemeine Botanik/Pflanzenökologie, University of Duisburg-Essen, Universitätsstraße 5, D-45117 Essen, Germany. E-mail: Heil_Martin@web.de

studying mechanisms that determine species-specific interactions (*11*). Whereas foraging ants are attracted by plant-derived food rewards in facultative interactions, obligate ant plants (myrmecophytes) house and nourish specialized ant colonies, which defend their hosts against herbivores, pathogens, and competing vegetation [reviewed in (*10*)]. Although both partners appear to be vitally dependent on such interactions, the mutualistic association even in the case of obligate myrmecophytes is transmitted horizontally: Both partners reproduce independently, and the mutualism has to be established anew in each subsequent generation (*12*). Such mutualisms are easily exploited by parasites or "cheaters" (*13*), and several parasites exploiting the plant-derived food resources have been described in ant/plant mutualisms (*14*–*16*). Filters excluding other species have therefore often been predicted (*10*). However, the existence and/or relevance of such filters for highly specific mutualisms such as ant/plant interactions have rarely been investigated (*17*, *18*). EFN appears to be an easily accessible target for exploitation by nonsymbiotic or even parasitic species, which, if they competed with the resident ants, would reduce the overall efficacy of the mutualistic interaction (*19*). We therefore assessed whether the EFNs secreted by related plant species having varying degrees of association with ants differ in their chemical compositions and whether those compositions are affected by any postsecretory mechanism.

We compared the EFN of four Central American *Acacia* myrmecophytes to the EFN of related yet nonmyrmecophytic species. Myrmecophytic *Acacia* species secrete EFN constitutively to nourish specialized mutualists, whereas the other species secrete EFN only in response to herbivore attack; this EFN attracts nonsymbiotic ants and thereby functions as an induced defense mechanism (*20*). Experiments were conducted at two different sites in Mexico. In our "cafeteria"-style experiments, we offered nectars of myrmecophytic and nonmyrmecophytic species, together with solutions of glucose, fructose, sucrose, and a mixture of these three sugars, to nonsymbiotic ants (species foraging in the vegetation and not regularly associated with *Acacia* myrmecophytes). Equal amounts of solutions at a standardized concentration were offered on the natural vegetation. All foraging ants arriving at the experimental site thus could freely choose among them. Ants feeding on the nectar droplets were counted repeatedly over at least 2 hours (*21*).

The species secreting the EFN was a significant source of variance in the ants' attendance [$P < 0.01$ according to one-way analysis of variance (ANOVA) for all four feeding experiments (Fig. 1)]. Nonsymbiotic ants (at least 11 species in total) significantly

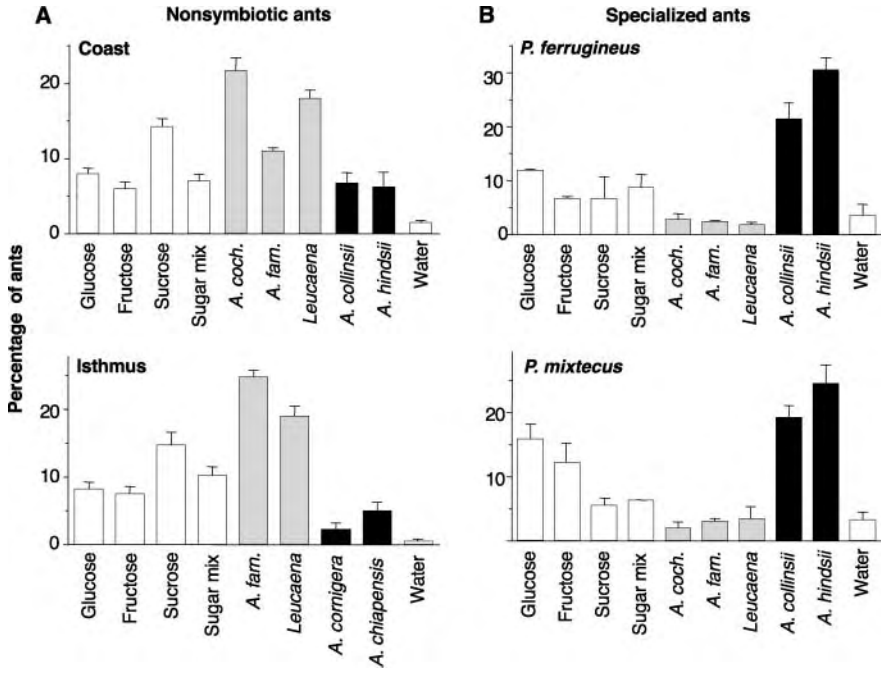


Fig. 1. Ant attendance to sugar solutions and EFNs secreted by certain *Acacia* species. Cafeteria-style experiments were conducted at two sites to test for attendance by the locally occurring nonsymbiotic ant species (**A**) (*21*). The same liquids were offered to two species (*P. ferrugineus* and *P. mixtecus*) specifically inhabiting *Acacia* myrmecophytes (**B**). Ant attendance is expressed as a percentage of all feeding ants that were attracted to one particular solution; bars represent means + SE. Nonmyrmecophytic species (*A. coch.*, *Acacia cochliacantha*; *A. farn.*, *Acacia farnesiana*; and *Leucaena*, *Leucaena leucocephala*) are shown in gray; myrmecophytes are shown in black. The type of sugar solution was a significant source of variance in the number of ants attracted to the different solutions ($P < 0.01$ in all four cases, one-way ANOVA on absolute total ant numbers, n = 4 days each for nonsymbiotic ants and $n = 5$ colonies each for specialized ants). Calculation of a priori contrasts revealed a significant ($P < 0.001$) difference between the EFN of myrmecophytes and the EFN of nonmyrmecophytes (see table S1 for all contrasts tested). See supporting online material for localization of the two study sites, coast and Isthmus.

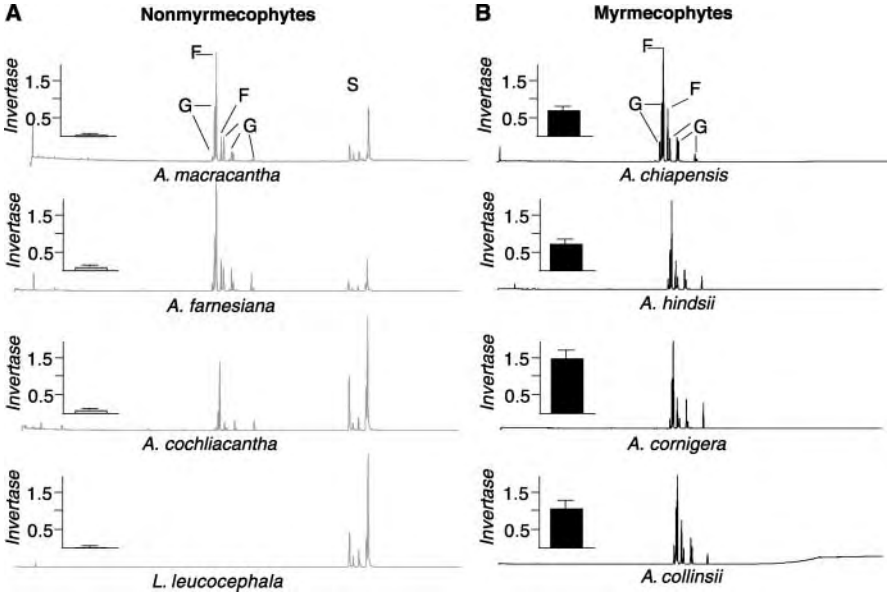


Fig. 2. Sugar profiles and invertase activity in EFNs. Chromatograms give relative abundances of sugars in the EFN of nonmyrmecophytic [(**A**), gray] and myrmecophytic [(**B**), black] species. Each sugar resulted in several peaks, which were identified by mass spectrometry (G, glucose; F, fructose; S, sucrose). Inserts at left: Invertase activity in the nectars is expressed as μg of glucose released per μl of EFN per minute, and bars represent means + SE. Species was a significant source of variance in invertase activity ($P < 0.01$ according to univariate ANOVA, five to six samples per species), and post hoc tests [least significant difference (LSD)] revealed that invertase activity in the EFNs of all myrmecophytes was significantly ($P < 0.05$) greater than in EFNs of all nonmyrmecophytes.

preferred the EFN secreted by nonmyrmecophytes to that of myrmecophytes (Fig. 1A). In contrast, two species of specialized *Acacia* ants (*Pseudomyrmex ferrugineus* and *P. mixtecus*) preferred the EFN of myrmecophytes to that of nonmyrmecophytes when they were subjected to a comparable experimental procedure (Fig. 1B).

The EFN secreted by myrmecophytes thus provides a valuable food source for the resident ants but is not attractive to nonsymbiotic EFN-feeding ants. Floral nectar or pollen can contain ant repellants in order to avoid ant/pollinator conflicts (*22*, *23*). Why is the EFN secreted by myrmecophytic *Acacia* species so unattractive to potential competitors of the resident ants? In the choice experiments, solutions of sucrose were much more attractive to nonsymbiotic ants than was the EFN of myrmecophytes (Fig. 1A). Sucrose, a common component of EFN, is generally very attractive to ants (*6*, *24*–*27*). Although amino acids can increase the attractiveness of EFN (*5*, *6*), sucrose thus was a likely candidate to cause the strong differences noted in our study. Gas chromatography–mass spectrometry analyses (*21*) revealed that the EFN of the nonmyrmecophytes always contained sucrose, as well as varying amounts of glucose and fructose (Fig. 2A). In contrast, the EFN of *Acacia* myrmecophytes contained only glucose and fructose (Fig. 2B). Adding sucrose to myrmecophyte EFN significantly increased its attractiveness to generalists but made it less attractive to specialized ants (Fig. 3). The lack of sucrose is thus one important factor (possibly among others) making EFN secreted by Central American *Acacia* myrmecophytes so unattractive to nonsymbiotic ants.

Is the absence of sucrose a consequence of a pre-secretion process, or do any postsecretory regulation mechanisms affect the carbohydrate composition of nectar? To study the underlying mechanism, we quantified invertase activity (encoded as EC 3.2.1.26 by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology, catalyzing the hydrolysis of sucrose into glucose and fructose) in the EFN of the study species (*21*). Invertase activity was detected in all samples of the myrmecophytic species we investigated but in only some samples of nonmyrmecophytic plants. On average, invertase activity ranged from 0.01 to 0.09 μg of glucose released μl^{-1} min^{-1} in the EFN of nonmyrmecophytes and 0.73 to 1.52 μg of glucose μl^{-1} min^{-1} in the EFN of myrmecophytes (inserts, Fig. 2). Although it does not exclude a selective secretion, this result shows that the EFN of the myrmecophytes is kept free of sucrose by postsecretory hydrolytic activity.

Because they are highly attractive to many different ant species (*6*, *24*–*26*), sucrose and other di- and trisaccharides appear to be a particularly important food component for ants. Why is the EFN with sucrose, then, less accepted by *Pseudomyrmex* ants that inhabit the *Acacia* myrmecophytes? Disaccharides and larger oligosaccharides must be cleaved into their monomers before they can be transported through membranes. Consequently, invertase activity has repeatedly been detected in ants (*25*, *28*–*30*). Invertase activity in extracts of the digestive system of the two specialized *Acacia* inhabitants (*21*) was significantly lower than in seven other species that live at the same site (Fig. 4). One of the latter species, *P. gracilis*, belongs to the same genus as the two *Acacia* mutualists but has only facultative interactions with *Acacia* myrmecophytes and can live independent from these plants. This species showed significantly greater invertase activity than did the specialized ants (Fig. 4).

Close reciprocal adaptations among *Acacia* myrmecophytes and the inhabiting ants *P.*

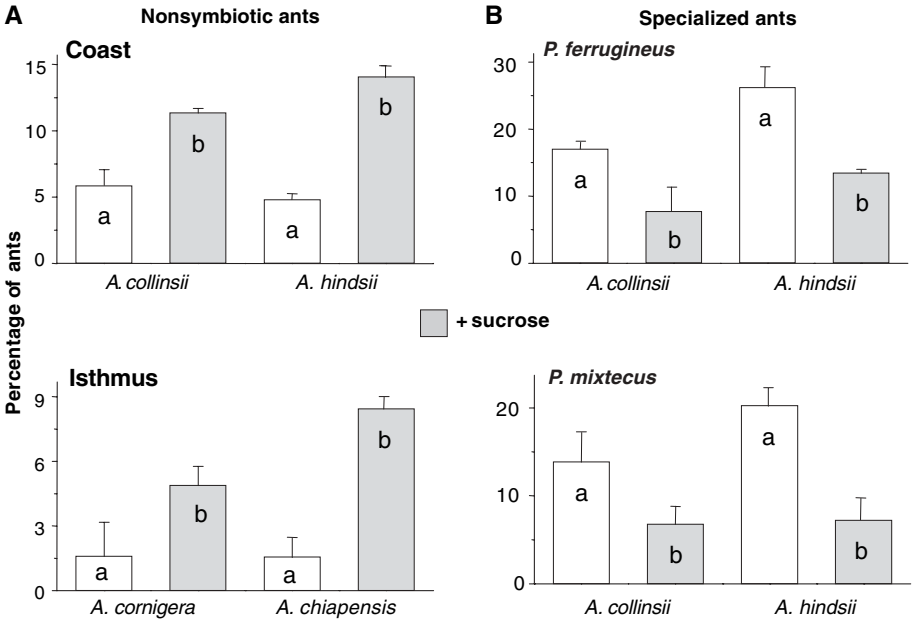


Fig. 3. Responses of ants to EFN nectar of two *Acacia* myrmecophytes (*A. cornigera* and *A. hindsii*) supplemented with sucrose. Responses of nonsymbiotic (**A**) and specialized (**B**) ants to nectars without (white bars) and with (gray bars) added sucrose [equal volumes of natural nectar and sucrose solution mixed, concentration 2% (weight/volume)] are given as percentages of ants attracted to these two solutions in cafeteria-style experiments. Nonsymbiotic ants were investigated at two sites (coast and isthmus), and the specialized *Acacia* inhabitants (*P. ferrugineus* and *P. mixtecus*) were investigated on *Acacia collinsii* shrubs. Bars represent means + SE. The numbers of ants attracted to nectar with and without added sucrose differed significantly [indicated by different letters; $P < 0.05$ according to paired *t* test, $n = 7$ feeding sites (A) or colonies (B) in all cases].

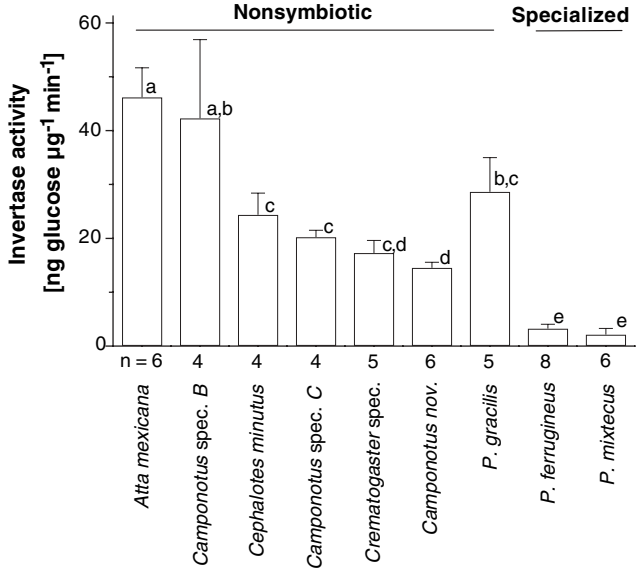


Fig. 4. Invertase activity in the digestive tracts of several ant species. Ants feeding on EFN and/or plant sap were collected from several colonies (sample numbers appear below bars representing means + SE) to compare their invertase activity to the activity in *Pseudomyrmex ferrugineus* and *P. mixtecus*, two species obligately inhabiting *Acacia* myrmecophytes. Species was a significant source of variance ($P < 0.001$, univariate ANOVA), and species labeled with different letters are significantly different ($P < 0.05$, LSD post hoc analysis). *Camponotus nov.*, *Camponotus novogranadensis*.

ferrugineus and *P. mixtecus* became apparent in this study. Invertase activity, which was very low in the mutualistic ants' digestive tracts, was significantly increased in the EFN that is secreted by these ants' host plants. Because myrmecophytism represents the derived trait among Central American *Acacia* species (*20*), the enzymatic activity involved in postsecretory regulation of nectar carbohydrate composition must have been greatly intensified during the evolution of these species' life history. This adaptation allows plants to present a valuable food source to their resident mutualists that will seldom be exploited by unspecialized competitors.

Few enzymes are known from any type of nectar, and those that have been characterized function in the antimicrobial defense of floral nectars (*31*–*33*). The invertase activity in the EFN of *Acacia* myrmecophytes, in contrast, affects the composition of the main nectar components: carbohydrates. To date, the source of this activity is not known, and even a microbial origin cannot be excluded. However, we detected free proteins in the EFN of myrmecophytes that are absent from the EFN of nonmyrmecophytes. The removal of these proteins from the EFN removes invertase activity as well, and data from preliminary matrix-assisted laser desorption/ionization–time-of-flight mass spectrometry characterization of those proteins are consistent with a putative plant origin (*34*). Similar patterns in invertase activity have been observed in plants growing in the field (where the EFN is immediately collected by ants) and in a greenhouse in Germany (where the EFN remains on nectaries for days). These observations support the idea that the invertase activity we observed results from an enzyme that is secreted by the plant.

The so-called filters that allow for specificity in horizontally transmitted mutualisms appear to be complex signals comprising elaborate blends of chemical compounds, and they often include cascades of consecutively activated genes. We found that even the presence (or absence) of a seemingly simple and common compound such as sucrose can be used as a filter to stabilize a specific symbiotic mutualism. The absence of sucrose from the EFN of myrmecophytes is an illustration of what had been called "adaptive specialization": an evolved trait that excludes less desirable partners in a multispecies association (*11*). Such simple compounds may have easily been overlooked in previous studies, if only because of their ubiquity, and should be considered in future studies.

References and Notes

1. H. G. Baker, I. Baker, in *Biochemical Aspects of Evolutionary Biology*, M. Nitecki, Ed. (Univ. of Chicago Press, Chicago, 1982), pp. 131–171.
2. B. Bentley, T. S. Elias, *The Biology of Nectaries* (Columbia Univ. Press, New York, 1983).
3. V. V. Roshchina, V. D. Roshchina, *The Excretory Function of Higher Plants* (Springer-Verlag, Berlin, 1993).
4. H. G. Baker, I. Baker, *Nature* **241**, 543 (1973).
5. J. Lanza, *Ecol. Entomol.* **16**, 203 (1991).
6. N. Blüthgen, K. Fiedler, *J. Anim. Ecol.* **73**, 155 (2004).
7. T. S. Elias, in *The Biology of Nectaries*, B. Bentley, T. S. Elias, Eds. (Columbia Univ. Press, New York, 1983), pp. 174–203.
8. J. D. Hoeksema, E. M. Bruna, *Oecologia* **125**, 321 (2000).
9. J. L. Bronstein, *Q. Rev. Biol.* **69**, 31 (1994).
10. M. Heil, D. McKey, *Annu. Rev. Ecol. Evol. Syst.* **34**, 425 (2003).
11. M. L. Stanton, *Am. Nat.* **162**, S10 (2003).
12. J. L. Bronstein, *Biotropica* **30**, 150 (1998).
13. D. M. Wilkinson, T. N. Sherratt, *Oikos* **92**, 377 (2001).
14. D. H. Janzen, *Science* **188**, 936 (1975).
15. D. K. Letourneau, *Science* **248**, 215 (1990).
16. N. E. Raine, N. Gammans, I. J. Macfadyen, G. K. Scrivner, G. N. Stone, *Ecol. Entomol.* **29**, 345 (2004).
17. J. L. Bronstein, *Ecol. Lett.* **4**, 277 (2001).
18. C. Brouat, N. Garcia, C. Andary, D. McKey, *Proc. R. Soc. London Ser. B* **268**, 2131 (2001).
19. M. Heil, A. Hilpert, R. Krüger, K. E. Linsenmair, *J. Trop. Ecol.* **20**, 201 (2004).
20. M. Heil *et al.*, *Nature* **430**, 205 (2004).
21. Materials and methods are available as supporting material on *Science* Online.
22. N. E. Raine, P. Willmer, G. N. Stone, *Ecology* **83**, 3086 (2002).
23. J. Ghazoul, *Ecol. Lett.* **4**, 295 (2001).
24. M. L. Cornelius, J. K. Grace, J. R. Yates, *Anz. Schaedlingskd. Pflanzenschutz Umweltschutz* **69**, 41 (1996).
25. J. L. Boevé, F. L. Wackers, *Oecologia* **136**, 508 (2003).
26. J. O. Stapel, A. M. Cortesero, C. M. DeMoraes, J. H. Tumlinson, W. J. Lewis, *Environ. Entomol.* **26**, 617 (1997).
27. R. K. Vandermeer, C. S. Lofgren, J. A. Seawright, *Fla. Entomol.* **78**, 144 (1995).
28. G. L. Ayre, *Can. Entomol.* **99**, 408 (1967).
29. B. L. Ricks, S. B. Vinson, *Entomol. Exp. Appl.* **15**, 329 (1972).
30. M. Heil, R. Büchler, W. Boland, *J. Chem. Ecol.* **30**, RC177 (2004).
31. W. J. Peumans, K. Smeets, K. van Nerum, F. van Leuven, E. J. M. van Damme, *Planta* **201**, 298 (1997).
32. R. W. Thornburg *et al.*, *Plant Syst. Evol.* **238**, 211 (2004).
33. C. Carter, R. W. Thornburg, *Trends Plant Sci.* **9**, 320 (2004).
34. A. Svatos, M. Heil, unpublished data.

35. We thank R. Krüger for enthusiastic support of field work; I. T. Baldwin, D. McKey, T. Hartmann, P. Feeny, E. Wheeler, and three anonymous referees for comments on earlier versions of this manuscript; and M. Verhaagh for determining ants. Financial support by the Deutsche Forschungsgesellschaft (DFG) (Emmy-Noether program and DFG grant He3169/3-1) and the Max-Planck-Society is gratefully acknowledged.

Supporting Online Material

www.sciencemag.org/cgi/content/full/308/5721/560/DC1
Materials and Methods
Table S1
References

15 November 2004; accepted 9 February 2005
10.1126/science.1107536

Retinoic Acid Controls the Bilateral Symmetry of Somite Formation in the Mouse Embryo

Julien Vermot,[1]*†‡ Jabier Gallego Llamas,[1]* Valérie Fraulob,[1] Karen Niederreither,[2] Pierre Chambon,[1] Pascal Dollé[1]‡

A striking characteristic of vertebrate embryos is their bilaterally symmetric body plan, which is particularly obvious at the level of the somites and their derivatives such as the vertebral column. Segmentation of the presomitic mesoderm must therefore be tightly coordinated along the left and right embryonic sides. We show that mutant mice defective for retinoic acid synthesis exhibit delayed somite formation on the right side. Asymmetric somite formation correlates with a left-right desynchronization of the segmentation clock oscillations. These data implicate retinoic acid as an endogenous signal that maintains the bilateral synchrony of mesoderm segmentation, and therefore controls bilateral symmetry, in vertebrate embryos.

The body plan of vertebrate embryos is overtly symmetric; only later in development do internal organs move into asymmetric positions. Among the most obviously symmetric embryonic structures are the left and right somitic columns, in which paired epithelial structures arise by segmentation of the paraxial mesoderm. Somite development relies on a "clock and wavefront" mechanism (*1*), in which a molecular oscillator that depends on Notch and Wnt pathways (the "segmentation clock") generates cyclic waves of gene expression along the presomitic mesoderm (PSM) (*2*, *3*). In addition, a caudal-to-rostral *Fgf8* (fibroblast growth factor 8) mRNA gradient acts as a moving wavefront (the "determination front"), triggering somite differentiation and setting the intersomitic boundaries (*4*). Retinoic acid (RA) plays multiple roles during patterning of the vertebrate anteroposterior axis. Altered RA

[1]Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/ULP/Collège de France, BP 10142, 67404 Illkirch Cedex, Strasbourg, France. [2]Departments of Medicine and Molecular Biology, Center for Cardiovascular Development, Baylor College of Medicine, Houston, TX 77030, USA.

*These authors contributed equally to this work.
†Present address: Beckman Institute, California Institute of Technology, Pasadena, CA 91125, USA.
‡To whom correspondence should be addressed. E-mail: dolle@igbmc.u-strasbg.fr (P.D.); julien@igbmc.u-strasbg.fr (J.V.)

signaling affects patterning of the vertebrae, generating homeotic transformations and/or segmentation defects (*5*). Experiments in chick and amphibian embryos have shown that RA signaling can counteract the *Fgf8* gradient in both the neural tube and PSM (*6*, *7*).

To define the role of RA signaling during mouse somitogenesis, we first used *lacZ* in situ hybridization (ISH) to characterize the pattern of activation of the RARE_hsp68_*lacZ* transgene, a sensitive reporter for the presence of endogenous RA (*8*). At somite stages 0 to 2, the transgene was active in both the somitic and PSM but not within the primitive streak (Fig. 1A) (*9*). From somite stages 4 to 10, RA activity was observed throughout the formed somites and within the rostral PSM up to a sharp, symmetric posterior boundary. Because this boundary was always located at the same distance from the last formed somite, the RA response appears to progress as a symmetric moving front during formation of the first somite pairs (Fig. 1, B to D).

To map the location of RA-responsive cells in embryos at somite stages 6 to 10, we combined immunofluorescent β-galactosidase (β-Gal) detection with ISH for *Mesp2*, whose rostral transcript boundary marks the border between presomites S-I and S-II [(*10*) for presomite nomenclature; (*11*)]. β-Gal$^+$ cells reached the *Mesp2* domain (Fig. 1, E to G), which shows that the RA response occurs in S-I. To determine whether the RA response is gradual, we examined embryos after increasing times of *lacZ* staining. No change was observed in the extent of the signal (Fig. 1, H to J); this result implies that cells respond to RA according to a sharp front rather than a gradient.

We mapped the location of the RA-responsive front with respect to the expression of the RA-metabolizing enzyme *Cyp26A1* (*12*). Both domains were consistently separated by nonlabeled PSM cells (Fig. 1, K and L), indicating that the RA-responsive front is not dictated by CYP26A1 activity. On the other hand, expression of the RA-synthesizing enzyme RALDH2 (retinaldehyde-specific dehydrogenase type 2) matched that of the reporter transgene in the rostral PSM (Fig. 1M). RALDH2 protein was no longer detected in the PSM from somite stages 10 to 12 (Fig. 1, N and O), although expression was later found in the mature somites (*9*). A similar drop in the activity of the RARE_*lacZ* transgene was seen in the PSM at these stages (fig. S1). Thus, a transition in RA biosynthesis occurs in the PSM during the formation of somites 11 and 12, which corresponds to the future cervicothoracic transition in the mouse.

To investigate the function of RA during somitogenesis, we analyzed *Raldh2*$^{-/-}$ embryos, which are known to exhibit abnormally small somites (*13*). The dynamics of somite formation was analyzed using *Uncx4.1* as a marker of mature somites. Until somite stage 7 or 8, somites were smaller in mutants but were evenly aligned along the left and right sides (Fig. 2, A and B). However, at later somite stages, some of the *Raldh2*$^{-/-}$ embryos exhibited asymmetric *Uncx4.1* patterns, such that at least one additional expression stripe was present on the left side (Fig. 2, C and D). Asymmetric patterns were also observed for *Meox1*, a pansomitic marker induced during somite formation in S-I, which indicates that the defect occurs within the somite-forming region (fig. S2).

Analysis of numerous *Uncx4.1*-hybridized *Raldh2*$^{-/-}$ embryos confirmed that asymmetric somite development first appears by somite stage 8 to 9 (Fig. 2F). The frequency of asymmetric patterns increased during the next cycles of somitogenesis, such that only 2 of 18 mutants exhibited a symmetric 11-somite pattern. In most mutants, one or more supernumerary somites were formed on the left side (Fig. 2F). Up to three additional somites could thus be seen (Fig. 2D). Eventually, a realignment of the last formed somites was seen in most mutants after formation of 15 or 16 somite pairs, although the adjacent somites were not properly aligned (Fig. 2E, bracket). One mutant showed arrested somite formation on the right side, with at least six somites fewer than on the left side (*9*).

We investigated the molecular basis of asymmetric somite development in *Raldh2*$^{-/-}$ embryos. The pace of somitogenesis is controlled by a molecular oscillator generating cyclic waves of gene expression within the PSM, in particular for genes of the Notch pathway (*3*, *4*). Among these, we analyzed *Lunatic fringe* (*Lfng*), whose cycling expression (*14*) can be visualized as three successive phases in which expression (i) is confined to the caudal PSM, (ii) "sweeps" along the PSM (Fig. 3A), and (iii) becomes localized to the forming somites (Fig. 3B). Asymmetric *Lfng* patterns were first observed in some *Raldh2*$^{-/-}$ embryos at somite stages 7 to 9, ranging from a subtle posterior shift on the right side (*9*) (n = 4 of 14) to a complete change in the oscillating phase patterns (Fig. 3C; n = 2 of 14). In slightly older mutants, "canonical" *Lfng* expression patterns were usually observed in the left PSM, whereas a supernumerary band of expression could be seen in the right PSM (Fig. 3D; n = 4 of 6). The right-side rostral bands were misaligned with the contralateral band,

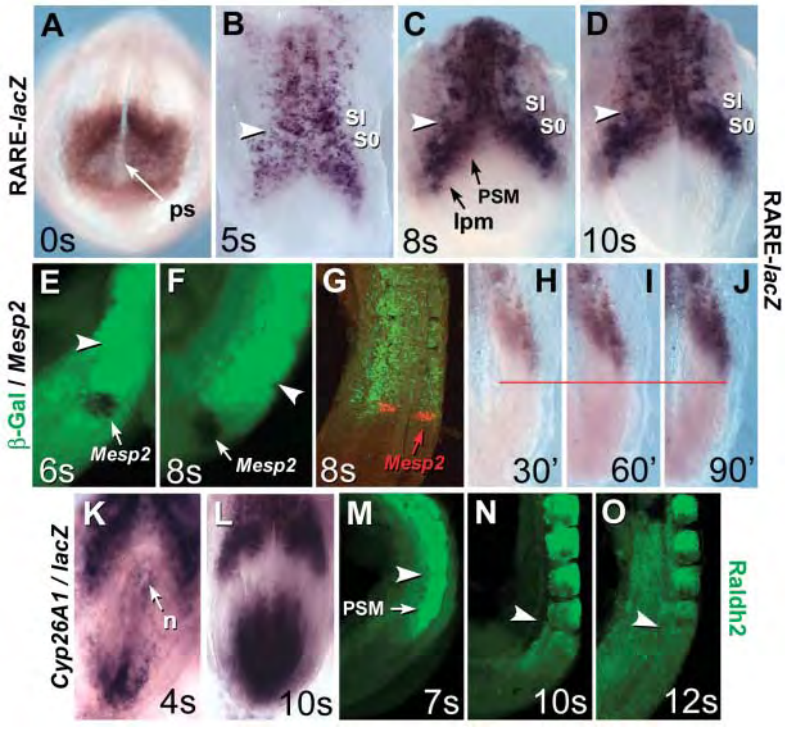


Fig. 1. Embryonic RA signaling is dynamically regulated during somitogenesis. (**A** to **D**) Whole-mount *lacZ* detection in RARE_hsp68_*lacZ* transgenic mouse embryos at successive somite stages (ventral views). lpm, lateral plate mesoderm; ps, primitive streak. (**E** to **G**) Combined detection of β-Gal (green) and *Mesp2* transcripts [black in (E) and (F), whole mounts; red in (G), confocal section] in RARE_hsp68_*lacZ* embryos. (**H** to **J**) *lacZ* signal in a transgenic embryo at somite stage 8 after increasing staining times (in minutes). Signal in the PSM demarcates a fixed boundary (red line). (**K** and **L**) Combined detection of *Cyp26A1* and *lacZ* transcripts in transgenic embryos (n, node). (**M** to **O**) RALDH2 protein immunodetection in wild-type embryos. Somite stages are as indicated (0s to 12s). White arrowheads point to the last formed intersomitic (SI/S0) boundary. Whatever the somite stage, SI and S0 designate the last formed mature (epithelialized) somite and the next forming somite, respectively [see (*10*)].

and a "salt and pepper" pattern of expressing and nonexpressing cells was seen between them (Fig. 3D) (fig. S3). At somite stages 15 and 16, *Lfng* expression was symmetric in the majority of *Raldh2* mutants (*9*) (n = 6 of 8). These results show that the progression of the waves of *Lfng* expression is no longer coordinated along the left and right PSM, during a temporal window correlating with asymmetric somite development, in the RA-deficient embryos.

We analyzed whether other "oscillating" genes have their expression patterns altered in *Raldh2*$^{-/-}$ mutants. The basic helix-loop-helix factor HES7 is a pivotal component of the molecular clock, as it represses *Lfng* in the PSM during each oscillation (*15*). Several *Raldh2*$^{-/-}$ embryos exhibited asymmetric *Hes7* expression patterns (Fig. 3, E and F; n = 4 of 10) that further suggested a delay in the progression of the oscillatory waves along the right-side PSM (Fig. 3F). Similar results were obtained for *Hes1* (n = 2 of 7) and for *Axin2* (n = 4 of 9), which encodes a repressor of Wnt signaling (*3*, *9*). Downstream genes such as *Mesp2* also exhibited asymmetric patterns (*16*). Thus, RA deficiency leads to a lack of left-right coordination of the waves of expression of various oscillating genes. Such a molecular defect has never been described in a mouse (or another vertebrate) mutant. It is reminiscent, however, of the misphased oscillatory patterns observed in wild-type zebrafish embryos exposed to a left-right temperature gradient, which increases the rate of somitogenesis on the warmer side (*17*).

The position of somite boundaries is controlled by a caudal-to-rostral *Fgf8* gradient that regresses posteriorly in concert with axis extension (*4*, *18*). We analyzed *Fgf8* expression in the *Raldh2*$^{-/-}$ embryos. At somite stages 4 to 8, the *Fgf8* expression domain was extended anteriorly in the mutant PSM (Fig. 4, A and B; n = 7 of 7). It therefore appears that an abnormal distribution of the *Fgf8* gradient may be linked to the compacted somite phenotype of *Raldh2* mutants, as reported for RA-deficient quail embryos (*6*). At somite stages 9 and 10, an ectopic distribution of *Fgf8* transcripts in the right PSM was also seen in some mutants (Fig. 4, C and D; n = 5 of 12). In the chick embryo, altering the *Fgf8* gradient symmetry by grafting an FGF8 bead on the

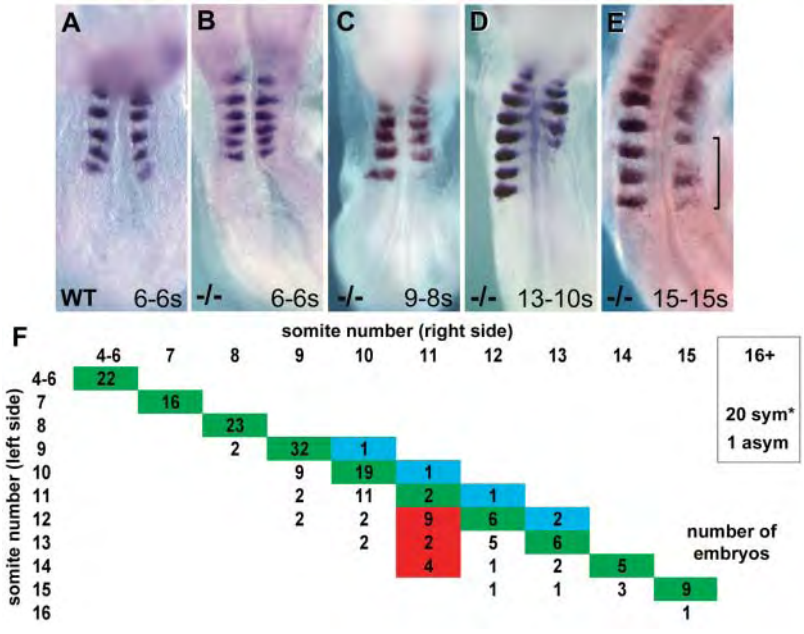
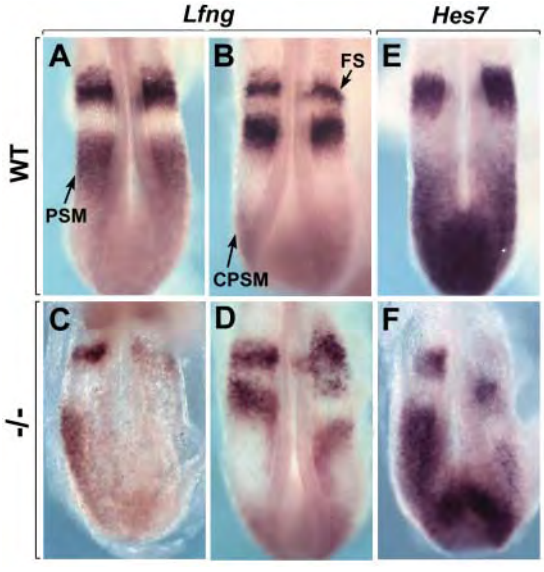


Fig. 2. Asymmetric somite development in *Raldh2*$^{-/-}$ embryos. (**A** to **E**) Whole-mount ISH of wild-type (A) and *Raldh2*$^{-/-}$ [(B) to (E)] embryos with an *Uncx4.1* probe (dorsal views). The numbers of formed somites on left and right sides are indicated below. (**F**) Summary of the numbers of *Raldh2*$^{-/-}$ embryos with equal (diagonal, green) or unequal somite numbers. The most acute phase of the asymmetry defect is shown in red. Few embryos exhibited an additional somite on the right side (blue). A "symmetric" (sym*) score at late somite stages (16 and above) refers to the location of the last formed somites, as more rostral somites were often misaligned [bracket in (E)].

Fig. 3. Whole-mount ISH of wild-type (**A**, **B**, and **E**) and *Raldh2*$^{-/-}$ (**C**, **D**, and **F**) embryos with *Lfng* [(A) to (D)] and *Hes7* [(E) and (F)] probes (dorsal views), showing abnormal progression of the molecular oscillations in *Raldh2*$^{-/-}$ embryos. PSM, presomitic mesoderm; CPSM, caudal presomitic mesoderm; FS, forming somite.



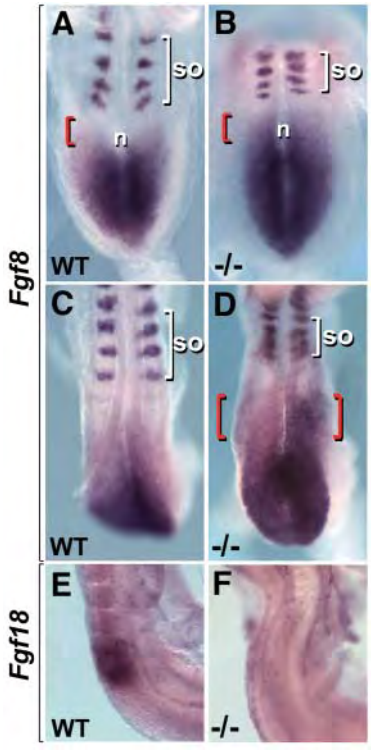


Fig. 4. Whole-mount ISH of wild-type (**A**, **C**, and **E**) and *Raldh2*$^{-/-}$ (**B**, **D**, and **F**) embryos with *Fgf8* + *Uncx4.1* [(A) to (D), dorsal views] and *Fgf18* [(E) and (F), lateral views] probes, showing altered *Fgf* expression in the somite-forming region of *Raldh2*$^{-/-}$ embryos. Red brackets in (A) and (B) are landmarks to compare *Fgf8* mRNA distribution (n, node; so, somites). Red brackets in (D) show unequal *Fgf8* levels along left and right PSM.

right side results in an asymmetric cyclic gene expression pattern and positioning of somite boundaries (*4*). This result was interpreted as a slowing of the FGF8 front posterior regression on the grafted side. Similarly, the unequal left-right *Fgf8* levels may reflect a delayed regression of the FGF8 front on the right side, contributing to the asymmetric positioning of somite boundaries in severely affected *Raldh2*$^{-/-}$ mutants. We analyzed whether expression of other *Fgf*s may also be altered. *Fgf18*, whose expression is normally restricted to SI and S0 (*19*) (Fig. 4E), was undetectable in *Raldh2*$^{-/-}$ embryos (Fig. 4F; $n = 6$ of 6). Hence, lack of FGF18 signaling may contribute to the somitic abnormalities and make the RA-deficient embryos particularly sensitive to changes in the *Fgf8* gradient.

Our results show that endogenous RA deficiency leads to a lack of coordination of somite formation between the left and right side of the mouse embryo, apparently due to a progressive delay of the oscillatory waves of gene expression in the right PSM. We have thus uncovered a genetically controlled mechanism that actively maintains the bilateral synchrony of mesoderm segmentation. Selective pressure for coordinated left-right somitogenesis is likely to occur throughout vertebrates, given that even a subtle lack of coordination would lead to defects in the bilateral symmetry of somitic derivatives (including the axial skeleton). Interestingly, asymmetric somite development occurs naturally in the cephalochordate *Amphioxus* (*20*), which suggests that the role of RA in the control of somitogenesis synchrony is an evolutionary acquisition of vertebrates.

The first waves of *Lfng* expression have been shown to be asymmetric when they reach Hensen's node in chick embryos (*21*). This biased expression is necessary for the left side–specific induction of *Nodal*, an early left-right determinant directly regulated by Notch signaling (*22*, *23*). Our data suggest that RA could act to synchronize the segmentation clock oscillations after the early phase of "natural," transient left-right asymmetry around the node region. Such an interaction with the left-right machinery is suggested by the lateralization of the somitogenesis defect, which almost always occurs on the same side of the RA-deficient embryos (*16*). Although the mechanisms involved in the breaking of left-right symmetry may differ between chick and mammal embryos (*24*), a common output is the generation and amplification of asymmetric signaling cascades in the mesoderm. As the RA signal progresses symmetrically along the left and right PSM of wild-type embryos (Fig. 1), it is likely to counteract the effect of left-right asymmetric signals in order to stabilize the progression of the molecular oscillations and thereby protect the bilateral synchrony of somite formation.

References and Notes

1. J. Cooke, E. C. Zeeman, *J. Theor. Biol.* **58**, 455 (1976).
2. I. Palmeirim, D. Henrique, D. Ish-Horowicz, O. Pourquié, *Cell* **91**, 639 (1997).
3. A. Aulehla *et al.*, *Dev. Cell* **4**, 395 (2003).
4. J. Dubrulle, M. J. McGrew, O. Pourquié, *Cell* **106**, 219 (2001).
5. M. Kessel, P. Gruss, *Cell* **67**, 89 (1991).
6. R. Diez del Corral *et al.*, *Neuron* **40**, 65 (2003).
7. T. A. Moreno, C. Kintner, *Dev. Cell* **6**, 205 (2004).
8. J. Rossant, R. Zirngibl, D. Cado, M. Shago, V. Giguère, *Genes Dev.* **5**, 1333 (1991).
9. J. Vermot *et al.*, data not shown.
10. O. Pourquié, P. P. Tam, *Dev. Cell* **1**, 619 (2001).
11. Y. Takahashi *et al.*, *Nat. Genet.* **25**, 390 (2000).
12. H. Fujii *et al.*, *EMBO J.* **16**, 4163 (1997).
13. K. Niederreither, V. Subbarayan, P. Dollé, P. Chambon, *Nat. Genet.* **21**, 444 (1999).
14. H. Forsberg, F. Crozet, N. A. Brown, *Curr. Biol.* **8**, 1027 (1998).
15. Y. Bessho *et al.*, *Genes Dev.* **15**, 2642 (2001).
16. See supporting data on *Science* Online.
17. Y. J. Jiang *et al.*, *Nature* **408**, 475 (2000).
18. A. Sawada *et al.*, *Development* **128**, 4873 (2001).
19. A. Liu *et al.*, *Development* **130**, 6175 (2003).
20. P. Cerfontaine, *Arch. Biol.*, T. 22 (1906).
21. A. Raya *et al.*, *Nature* **427**, 121 (2004).
22. L. T. Krebs *et al.*, *Genes Dev.* **17**, 1207 (2003).
23. A. Raya *et al.*, *Genes Dev.* **17**, 1213 (2003).
24. J. Capdevila, K. J. Vogan, C. J. Tabin, J. C. Izpisùa Belmonte, *Cell* **101**, 9 (2000).

25. We thank B. Schuhbaur for technical help, J. Rossant for RARE-*lacZ* mice, P. McCaffery for antibody to RALDH2, and K. Storey and O. Pourquié for discussions. Supported by funds from CNRS, INSERM, Collège de France, Ministère de la Recherche, and Institut Universitaire de France. J.V. was supported by fellowships from the Association pour la Recherche sur le Cancer and Fondation pour la Recherche Médicale.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1108363/DC1
Materials and Methods
SOM Text
Figs. S1 to S5

6 December 2004; accepted 9 February 2005
Published online 24 February 2005;
10.1126/science.1108363
Include this information when citing this paper.

Image Cytometer

The Cell Lab IC 100 Image Cytometer is a flexible, automated tool for quantitative cellular analysis. Based on unique imaging technology, it enables the identification of cellular events such as signaling cascades, translocation, protein tracking and migration, and surface and receptor activation. With the system, processes can be studied both spatially and temporally, in the context of the living cell. The software enables researchers to design novel assays and specify customized measurements. Suitable as a basic research tool for the academic core facility or for drug discovery work, the IC 100 can be used for a range of applications, including small-interfering RNA, protein expression, nuclear translocation, and toxicology. The information it gathers can be used to develop cell-based assays that identify and validate potential disease targets and to measure cellular response to potential therapeutics.

Beckman Coulter For information 800-742-2345 www.beckmancoulter.com

Autofluorescence Eliminator Reagent

With increasing age, the autofluorescent pigment lipofuscin accumulates in the cytoplasm of many cell types, including neurons. The presence of lipofuscin granules can complicate the use of fluorescence microscopy in the central nervous system because of its broad excitation and emission spectra, which overlap those of commonly used fluorophores. The Autofluorescence Eliminator Reagent can reduce or eliminate lipofuscin-like autofluorescence without adversely affecting other fluorescent label in sections of human, monkey, and rat neural tissue as well as other tissues.

Chemicon International For information 800-437-7500 www.chemicon.com

Formalin-Fixed Tissue Analysis

The Liquid Tissue MS Protein Prep Kit makes possible mass spectrometry (MS) analysis of formalin-fixed tissue. This simple biochemical protocol enables discovery and validation of protein biomarkers in the large number of archived tissue collections worldwide with well-documented clinical outcomes, including disease progression, drug response, and toxicity. These collections have been an untapped resource for proteomic analysis. The kit allows users to extract total protein from formalin-fixed, paraffin-embedded tissue for mass spectrometry analyses across a number of different MS platforms. Results compare favorably with results reported with other biological samples.

Expression Pathology
For information 301-977-3654
www.expressionpathology.com

For more information visit GetInfo, *Science*'s new online product index at *http://science.labvelocity.com*

From the pages of GetInfo, you can:
- Quickly find and request free information on products and services found in the pages of *Science*.
- Ask vendors to contact you with more information.
- Link directly to vendors' Web sites.

Electrophoresis Hydrogel

Blot-EX is a new nonacrylamide, nontoxic, high-resolution electrophoresis precast hydrogel designed to increase protein transfer efficiency during protein immunoblotting. Blot-EX offers overall recovery efficiency of up to 90%, obtained with significantly reduced transfer times (less than 10 min) compared with industry-standard polyacrylamide precast gels. Blot-EX operates in the size range of 30 to 200 KDa. The improvement in protein transfer is more important in the upper part of the ranges, which corresponds to enzymes and receptors.

Elchrom Scientific For information +41 41 747 25 41 www.elchrom.com

Multi-Purpose Image Analyzer

The ChemiGeniusXE 3 multi-purpose image analysis system is designed for laboratories that need a system for rapidly imaging in chemiluminescent, ultraviolet, and white-light applications. Inside a computer-controlled darkroom with auto-lock sliding doors, the ChemiGeniusXE 3 makes use of a 3.2-million pixel resolution, 16-bit charge-coupled device camera for unrivalled separation of images with close bands or spots. The camera, offering 90% quantum efficiency, can detect a range of colored dyes. It is cooled to –60°C to reduce noise when developing chemiluminescent signals over long exposure times. Both these features allow scientists to capture faint signals from any commercially available dyes, guaranteeing the most accurate sample quantification. This versatile system has interchangeable and fixed lenses and an adjustable multi-height sample tray so users can work with both large and small gels and blots. It comes complete with a large (25 cm × 30 cm), evenly lit transilluminator, white light pad, integral epi-ultraviolet (254 nm and 365 nm) illumination, and a computer-controlled, 10-position filter wheel for imaging different chemiluminescent and fluorescent dyes.

Syngene For information 800-686-4407 www.syngene.com

Literature

GeoScienceWorld (GSW) has launched a new portal containing 30 leading science journals plus GeoRef. The journals include nine not previously available on the Internet. The many advantages geoscientists will find at the site include references being hyperlinked among the GSW journals and other HighWire Press journals, GeoRef, ISI's Web of Science, CrossRef, and Medline. There is access to the current edition of GeoRef, the comprehensive database of geoscience literature, with weekly updates scheduled. Searches can be performed simultaneously among all GSW journals and all of GeoRef. Maps and images can be looked up on the Alexandria Digital Library. Users can download bibliographic references to various citation managers, including file formats for EndNote, Reference Manager, ProCite, BibTex, and Medlars.

GeoScienceWorld For information 703-671-4791 www.geoscienceworld.org

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and government organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier by visiting www.science.labvelocity.com on the Web, where you can request that the information be sent to you by e-mail, fax, mail, or telephone.